



**HAL**  
open science

# Identification et caractérisation de la famille de protéines à domaine LIM chez le peuplier

Dominique Arnaud

► **To cite this version:**

Dominique Arnaud. Identification et caractérisation de la famille de protéines à domaine LIM chez le peuplier. Sciences du Vivant [q-bio]. Université d'Orléans, 2007. Français. NNT: . tel-02824561

**HAL Id: tel-02824561**

**<https://hal.inrae.fr/tel-02824561>**

Submitted on 6 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE D'ORLEANS



**THESE PRESENTEE A L'UNIVERSITE D'ORLEANS  
POUR OBTENIR LE GRADE DE  
DOCTEUR DE L'UNIVERSITE D'ORLEANS**

**Discipline : Physiologie Génétique Moléculaire**

**PAR**

**ARNAUD Dominique**

**Identification et caractérisation de la famille de protéines à  
domaine LIM chez le peuplier**

Soutenue le : **Novembre 2007**

**MEMBRES DU JURY :**

- |   |                             |
|---|-----------------------------|
| - <b>Gilles PILATE</b> , Directeur de recherche, INRA Orléans   | <b>Directeur de thèse</b>   |
| - <b>Daniel LOCKER</b> , Professeur, CNRS Orléans   | <b>Codirecteur de thèse</b> |
| - <b>André STEINMETZ</b> , Directeur de recherche, Centre de Recherche<br>Public Santé, Luxembourg        | <b>Rapporteur</b>           |
| - <b>Jean-Louis JULIEN</b> , Professeur, Université Blaise Pascal,<br>Clermont Ferrand                    | <b>Rapporteur</b>           |
| - <b>Lise JOUANIN</b> , Directeur de recherche CNRS, INRA Versailles                                      | <b>Examineur</b>            |
| - <b>Eric LAINE</b> , Professeur, Université d'Orléans, Antenne Scientifique<br>Universitaire de Chartres | <b>Examineur</b>            |



## Avant Propos

Je remercie chaleureusement **André Steinmetz** et **Jean-Louis Julien** d'avoir accepté d'être mes rapporteurs de thèse, pour leur investissement et leur intérêt pour mon travail.

Je tiens aussi à remercier **Lise Jouanin** et **Eric Lainé** en tant qu'examineurs, et pour le jugement qu'ils porteront sur ma thèse.

Je remercie également **Daniel Locker** d'avoir accepté d'être mon co-directeur de thèse, et pour ses conseils pertinents qui ont contribué à la réalisation de ce travail.

J'adresse ma reconnaissance à **Gilles Pilate** pour m'avoir accueilli à l'INRA d'Orléans, pour avoir été mon directeur de thèse et pour son encadrement.

Merci **Annabelle** pour ta gentillesse, ta patience, ton aide et tout particulièrement pour ton soutien sans failles lors des dernières phases de la thèse. **Jean-Charles**, merci pour tes points de discussion « titilleux » mais toujours très justes, et aussi pour ta disponibilité et tes conseils de manips qui m'ont beaucoup apporté.

Je voudrais aussi remercier tous ceux du CBM qui m'ont soutenu dans ce travail et offert leurs services. Tout d'abord, **Martine Decoville** pour sa participation à mes comités de thèse, ses bons conseils et sa patience lors des manips ChIP. Merci **Hélène Bénédetti** d'avoir aussi participé à mes comités de thèse, de m'avoir accueilli dans ton équipe si sympathique et pour ton encadrement efficace lors des manips de purif de cette satanée protéine « chinoise ». Enfin, merci à **Alain Roussel** et **Bertrand Castaing**, vous qui avez pris de votre temps pour m'expliquer avec passion la biochimie des protéines, et aussi **Franck Coste** pour son savoir faire dans la purif de protéines.

Merci à **Jean-Charles Bastien** qui m'a accueilli et m'a permis de mettre en œuvre mon travail de recherche.

Je tiens aussi à remercier **Pascal Gantet** pour sa participation à mes comités de thèse. Dommage, qu'on n'ait pas eu le temps de discuter un peu plus de mes manips.

Merci **Marie-Claude** pour ton efficacité dans le soutien logistique de mes manips, et surtout pour toutes ces séquences d'ADN ! Merci à vous **Nathalie** et **Nadège** pour votre gentillesse, et assistance à tout moment en bioch ou en CIV. **Françoise** et **Alain** pour leurs aides en microscopie. Merci également **Annie** et **Gérald** pour votre dévouement. Je remercie aussi **Isabelle** pour ton aide précieuse dans l'analyse des séquences, **Guy** et **Eric** pour votre aide en informatique. Je n'oublie pas **Franck** et **Karine** pour m'avoir accompagné dans la recherche de biblio. Merci **Marc** et **Christophe** de m'avoir fait découvrir les fleurs de peupliers et l'apport de matériels contribuant à l'originalité de la thèse. Merci **Philippe** et **Christian** pour vos conseils et esprit critique tout au long de ma thèse. Merci aussi à vous **Corinne**, **Vanina** et **Véronique** pour votre aide à la paillasse. Merci **Jean-Paul** pour ces soirées sympas au mur d'escalade de la fac. Enfin, merci **Miyuki** d'avoir été ma voisine de bureau et pour ton soutien moral.

En souvenir à **Rozenn** et **Romain** qui nous ont quitté beaucoup trop tôt.

Aussi merci à toi ma **Nadia** qui m'a compris, aidé et encouragé pendant cette thèse. Je t'embrasse très fort !

Enfin, je finirais pas ceux auxquels je tiens le plus, mes **parents** et mon **frère**. Merci pour votre soutien chacun à votre manière.



# SOMMAIRE

<b>Introduction Générale</b>	<b>1</b>
<b>1 Le peuplier</b>	<b>1</b>
1.1. <i>Importance économique en France</i>	1
1.2. <i>Les peupliers appartiennent à la famille des Salicaceae et au genre Populus</i>	3
1.3. <i>Développement et morphologie du peuplier</i>	5
1.3.1. Morphologie et développement des inflorescences mâles	7
1.3.2. Morphologie et développement des inflorescences femelles	8
1.3.3. Fécondation et formation du tube pollinique	10
1.3.4. Morphologie et développement des organes fructifères	11
1.4. <i>Pourquoi avoir choisi le peuplier (Populus trichocarpa) comme arbre modèle ?</i>	13
1.5. <i>Le génome séquencé de Populus trichocarpa</i>	15
<b>2 Aspects physiologique et moléculaire de la formation des tissus vasculaires primaire et secondaire chez les plantes et le peuplier</b>	<b>19</b>
2.1. <i>La formation du système vasculaire primaire</i>	19
2.2. <i>La formation du bois ou xylogénèse chez le peuplier</i>	21
2.2.1. La zone cambiale, une région méristématique de division et d'expansion cellulaire	22
2.2.2. L'expansion cellulaire et l'élaboration de la paroi cellulaire primaire	25
2.2.3. La formation de la paroi secondaire	27
2.2.4. L'étape finale : la mort cellulaire programmée	30
2.3. <i>Rôle du cytosquelette dans la formation du bois</i>	30
2.3.1. Les microtubules	31
2.3.2. Les microfilaments d'actine	34
2.4. <i>La biosynthèse de cellulose et les cellulose synthases chez les plantes</i>	38
2.4.1. La biosynthèse de cellulose	38
2.4.2. La famille et la structure des cellulose synthases (CesA) de plante	39
2.4.3. Les cellulose synthases d'Arabidopsis thaliana	41
2.4.4. Les cellulose synthases de peuplier	43
2.5. <i>La lignine, un composé majeur des plantes et important lors de la formation du bois</i>	44
2.5.1. Lignification de la paroi cellulaire secondaire	45
2.5.2. La composition des polymères de lignine	45
2.5.3. La voie de biosynthèse de la lignine	46
2.5.4. Les gènes codant pour les enzymes de la voie de biosynthèse des lignines	48
2.5.5. Les phénylalanine ammonia lyases (PAL)	49
2.5.6. Les 4-coumarate:CoA ligases (4CL)	50
2.5.7. Les cinnamyl alcool déhydrogénases (CAD)	52
2.5.8. Régulation transcriptionnelle des gènes de la voie de biosynthèse des lignines	54
2.6. <i>Les facteurs de transcription impliqués dans la formation de la paroi secondaire et la biosynthèse des lignines</i>	56
2.6.1. Les facteurs de transcription de type NAC et la formation de la paroi secondaire des fibres	56
2.6.2. Les facteurs de transcription MYB impliqués dans la lignification	58
<b>3 Le bois de tension, bois de réaction chez les feuillus</b>	<b>60</b>
3.1. <i>Stimuli induisant la formation de bois de réaction</i>	60
3.2. <i>La différenciation du bois engendre des contraintes de maturation dans les tiges</i>	61
3.3. <i>Le bois de compression chez les conifères</i>	63

3.4.	<i>Caractéristiques morphologiques du bois de tension</i>	64
3.5.	<i>La paroi des fibres de bois tendu est caractérisée par la présence d'une couche gélatineuse chez le peuplier</i>	66
3.6.	<i>Le bois de tension contient moins de lignine</i>	67
3.7.	<i>Organisation du cytosquelette dans les cellules de bois de tension</i>	68
3.8.	<i>Le bois de tension, facilement inductible est un bon modèle d'étude</i>	69
3.9.	<i>Modification de l'expression des gènes dans le bois de tension</i>	70
3.9.1.	Les protéines fascioline-like à arabinogalactanes (FLAs) seraient des protéines pariétales essentielles à la formation de bois de tension	71
3.9.2.	Induction dans le bois de tension de l'expression des gènes impliqués dans la biosynthèse de cellulose	72
3.9.3.	Expression des gènes impliqués dans l'organisation du cytosquelette dans le bois de tension	74
3.9.4.	Les gènes impliqués dans la synthèse de la lignine	77
3.9.5.	Régulation de l'expression des gènes dans le bois de tension, les facteurs de transcription	78
<b>4</b>	<b>Les protéines à domaine LIM</b>	<b>80</b>
4.1.	<i>Les protéines à domaine LIM chez les animaux</i>	80
4.1.1.	Séquence et structure des domaines LIM	80
4.1.2.	Les domaines LIM sont des modules d'interaction protéine/protéine et peuvent fixer l'ADN	82
4.1.3.	Les protéines contenant des domaines LIM chez les eucaryotes	83
4.1.4.	Les protéines zyxine et Hic5 sont associées aux adhésions focales et fibres de stress d'actine	86
4.2.	<i>La famille de protéines riches en cystéines (CRP) chez les eucaryotes</i>	87
4.2.1.	Structure des protéines CRP	87
4.2.2.	Les gènes CRP sont préférentiellement exprimés dans les cellules musculaires	89
4.2.3.	Les protéines CRP ont une double localisation nucléaire et cytoplasmique	90
4.2.4.	Dans le noyau, les protéines CRP interagissent avec des facteurs de transcription	90
4.2.5.	Les protéines CRP cytoplasmiques sont associées aux filaments d'actine	92
4.3.	<i>Les protéines à domaine LIM chez les plantes</i>	95
4.3.1.	Structure des protéines LIM chez les plantes	96
4.3.2.	Les spécificités structurales des protéines à domaine LIM de plantes sont corrélées avec des profils d'expression distincts	97
4.3.3.	Les protéines LIM subissent des modifications post-traductionnelle	101
4.4.	<i>Les protéines LIM de plante ont une double localisation subcellulaire</i>	102
4.4.1.	Dans le noyau, la protéine NtWLIM1 agit comme un facteur de transcription impliqué dans la voie de biosynthèse des phénylpropanoïdes	103
4.4.2.	Dans le cytoplasme, la protéine NtWLIM1 se lie et empaquette les filaments d'actine	105
<b>5</b>	<b>Objectifs de la thèse et démarche expérimentale</b>	<b>108</b>

## **Partie 1. Analyse génomique et phylogénétique de la famille de protéines à domaine LIM chez le peuplier et les plantes** **109**

### **1 Matériel et méthodes** **109**

- |      |  |     |
|------|--|-----|
| 1.1. | <i>Caractérisation des séquences d'ADNc et de gènes LIM de peuplier</i>                      | 109 |
| 1.2. | <i>Recherche de séquences codant pour des protéines à deux domaines LIM chez les plantes</i> | 111 |
| 1.3. | <i>Analyses de séquences et analyses phylogénétiques des protéines à domaine LIM</i>         | 112 |

### **2 Résultats** **113**

2.1.	<i>Isolation des ADNc codant pour des protéines à domaine LIM chez le peuplier</i>	113
2.2.	<i>Le génome de Populus trichocarpa contient douze modèles de gène LIM.</i>	116
2.3.	<i>Identification de polymorphisme de séquences nucléotidique et protéique au sein des espèces de peuplier</i>	120
2.4.	<i>Les génomes d'Arabidopsis et de riz contiennent six modèles de gènes LIM</i>	122
2.5.	<i>Organisation génomique des gènes LIM chez le peuplier, Arabidopsis et le riz</i>	125
2.6.	<i>Identification d'étiquettes de séquences exprimées (ESTs) homologues aux protéines à domaine LIM chez les plantes</i>	129
2.7.	<i>Analyse phylogénétique des protéines à domaine LIM de plante</i>	131
2.7.1.	Quatre groupes différents $\alpha$ LIM1, $\beta$ LIM1, $\gamma$ LIM2, et $\delta$ LIM2	131
2.7.2.	Le groupe $\alpha$ LIM1	135
2.7.3.	Le nouveau groupe $\beta$ LIM1	140
2.7.4.	Le groupe $\gamma$ LIM2	140
2.7.5.	Le groupe $\delta$ LIM2	143
2.8.	<i>Analyse de séquences et de la structure des protéines à domaine LIM de plante</i>	145
2.8.1.	Comparaison des séquences des domaines LIM entre les protéines CRP animales et les protéines à domaine LIM de plante	145
2.8.2.	Spécificités structurales des protéines à domaines LIM au sein de chaque sous-groupe phylogénétique et identification de sites putatif de phosphorylation et glycosylation	148
<b>3</b>	<b>Discussion</b>	<b>150</b>
3.1.	<i>Les SNPs observés dans les gènes LIM de peuplier peuvent avoir une influence sur la fonction des protéines</i>	150
3.2.	<i>Les gènes LIM chez le peuplier apparaissent tous dupliqués</i>	151
3.3.	<i>Les gènes LIM de plantes ont une structure intron/exon très conservée</i>	152
3.4.	<i>L'expression des gènes WLIM2 peut être régulée par un mécanisme d'épissage alternatif</i>	153
3.5.	<i>Les résultats d'analyses phylogénétiques sont corrélés avec les données d'expression</i>	154
3.5.1.	Le groupe $\alpha$ LIM1 hétérogène, comporte des sous-groupes avec des spécificités d'expression.	154
3.5.2.	Les protéines LIM1 de conifères et de fougères sont ancestrales aux groupes $\alpha$ LIM1 et $\beta$ LIM1	158
3.5.3.	Le nouveau groupe $\beta$ LIM1 est encore à explorer	160
3.5.4.	Les gènes du groupe $\gamma$ LIM2 important pour la plante seraient probablement régulés par des mécanismes ressemblant à de l'épigénétique	161
3.5.5.	Les gènes du groupe $\delta$ LIM2 sont préférentiellement exprimés dans le pollen	162
3.6.	<i>Vers une nouvelle caractérisation des séquences consensus des deux domaines LIM chez les plantes</i>	164
3.7.	<i>Les variations structurelles des protéines à domaine LIM peuvent refléter de différences fonctionnelles</i>	166
<b>4</b>	<b>Conclusions</b>	<b>168</b>

## **Partie 2. Approche de l'étude fonctionnelle de la famille LIM chez le peuplier par une analyse d'expression de gènes** **169**

### **1 Matériel et méthodes** **169**

1.1.	<i>Matériel</i>	169
1.1.1.	Echantillonnage des organes végétatifs	169



	1.1.2.	Protocole de mise en tension des arbres et prélèvements des bois tendu et opposé	171
	1.1.3.	Echantillonnage des organes reproducteurs	172
1.2.	<i>Méthodes</i>		173
	1.2.1.	Analyse de la distribution des ESTs codant pour des protéines à domaine LIM chez le peuplier	173
	1.2.2.	Extraction des ARN totaux de peuplier	174
	1.2.3.	Quantification des ARN totaux	174
	1.2.4.	Reverse transcription des ARNm en ADNc	175
	1.2.5.	Caractéristiques du choix des amorces pour les RT-PCR	175
	1.2.6.	Test des amorces sur vecteur et mise au point des conditions PCR	177
	1.2.7.	Réaction de polymérisation en chaîne « Polymerase chain reaction » (PCR)	177
	1.2.8.	Séquençage des produits RT-PCR	178
	1.2.9.	Etude de l'expression par Northern Blot du gène PtGLIM1a	178
<b>2</b>	<b>Résultats</b>		<b>179</b>
	2.1.	<i>Analyse de la distribution des ESTs codant pour des protéines à domaine LIM chez le peuplier</i>	179
	2.2.	<i>Etude de l'expression des gènes LIM de peuplier par RT-PCR semi quantitative</i>	183
	2.2.1.	Analyse des profils d'expression des gènes PtLIM dans les tissus végétatifs de peupliers.	183
	2.2.2.	Analyse de la quantité de transcrits PtLIM dans les bois tendu et opposé prélevés sur de jeunes arbres	188
	2.2.3.	Analyse de la quantité de transcrits PtLIM dans les organes reproducteurs du peuplier	191
	2.3.	<i>L'analyse en Northern blot confirme l'induction de l'expression du gène PtGLIM1a dans le bois de tension</i>	194
<b>3</b>	<b>Discussion</b>		<b>195</b>
	3.1.	<i>L'analyse de la répartition des ESTs révèle la spécificité d'expression vasculaire des gènes LIM chez le peuplier.</i>	195
	3.2.	<i>L'étude de l'expression des gènes PtLIM par RT-PCR semi quantitative révèle des spécificités d'expression tissulaire particulières.</i>	196
	3.2.1.	Corrélations entre les résultats d'analyse de distribution d'EST et les résultats RT-PCR	197
	3.2.2.	L'expression des gènes PtLIM est régulée en fonction des différentes étapes de différenciation du xylème et du phloème	198
	3.2.3.	Les gènes PtLIM sont également exprimés dans les tissus vasculaires de la racine	200
	3.2.4.	Les gènes PtLIM joueraient un rôle dans la formation du bois de tension	201
	3.2.5.	Les gènes PtPLIM2a et b et Pt $\delta$ LIM2a et b sont spécifiquement exprimés dans le pollen	204
	3.2.6.	Les gènes PtLIM interviennent dans la formation des fibres de coton chez le peuplier	206
	3.2.7.	Les gènes dupliqués chez le peuplier	208
<b>4</b>	<b>Conclusions</b>		<b>210</b>
<b>Partie 3. Approche biochimique de l'étude de la protéine PtaGLIM1a</b>			<b>211</b>
<b>1</b>	<b>Matériels et méthodes</b>		<b>211</b>
	1.1.	<i>Souches de bactéries utilisées</i>	211
	1.2.	<i>Préparation de bactéries compétentes pour la transformation par choc thermique</i>	212

1.3.	<i>Transformation par choc thermique des bactéries compétentes</i>	212
1.4.	<i>Principe du clonage de gènes en utilisant le système Gateway®</i>	213
1.5.	<i>Clonage de l'ADNc PtaGLIM1a dans le vecteur pENTR™/D-TOPO® avec le système Gateway®</i>	214
1.6.	<i>Recombinaison du gène PtaGLIM1a dans les vecteurs de destination pDEST™15 et pDEST™17 avec le système Gateway®</i>	215
1.7.	<i>Production de la protéine recombinante GST-PtaGLIM1a</i>	216
1.7.1.	Mise au point des conditions de culture des bactéries pour la production de protéine recombinante GST-PtaGLIM1a	216
1.7.2.	Mise au point de la solubilisation des protéines recombinantes GST-PtaGLIM1a	217
1.7.3.	Production et purification sur de la protéine recombinante GST-PtaGLIM1a	218
1.7.4.	Clivage à la thrombine de la protéine GST-PtaGLIM1a et capture de la thrombine par la streptavidine	219
1.7.5.	Élimination de la GST sur colonne Glutathione Sepharose 4B :	219
1.7.6.	Test d'élimination des protéines contaminantes par tamisage moléculaire sur une colonne de gel filtration.	220
1.7.7.	Test d'élimination des protéines chaperonnes GroEL	220
1.8.	<i>Production et purification de la protéine recombinante 6His-PtaGLIM1a</i>	221
1.8.1.	Test de la production et de la solubilisation de la protéine 6His-PtaGLIM1a dans différentes bactéries	221
1.8.2.	Purification de la protéine recombinante 6His-PtaGLIM1a sous forme de corps d'inclusion puis par affinité sur une résine de Cobalt	221
1.9.	<i>Production d'anticorps polyclonaux anti-PtaGLIM1a :</i>	223
1.10.	<i>Extraction de protéines totales de peuplier</i>	224
1.11.	<i>Dosage de la quantité de protéine par la méthode de Bradford (Bradford, 1976)</i>	225
1.12.	<i>Migration des protéines en condition dénaturante en gel SDS-PAGE (Laemmli, 1970)</i>	225
1.13.	<i>Coloration au bleu de Coomassie :</i>	226
1.14.	<i>Transfert des protéines sur membrane de nitrocellulose et réalisation de Dot-blot avec le peptide P-GLIM1a</i>	226
1.15.	<i>Western blot</i>	227
<b>2</b>	<b>Résultats</b>	<b>228</b>
2.1.	<i>Production de protéines recombinante GST-PtaGLIM1a :</i>	229
2.1.1.	Mise au point de la production et de la solubilisation de la protéine recombinante GST-PtaGLIM1a	229
2.1.2.	Purification de la protéine recombinante GST-PtaGLIM1a par affinité sur résine de Glutathione Sépharose 4B et clivage de la GST par la thrombine	232
2.1.3.	Test de l'élimination des protéines contaminantes co-purifiées avec la protéine GST-PtaGLIM1a	234
2.2.	<i>Production de protéines recombinante 6His-PtaGLIM1a :</i>	235
2.2.1.	Mise au point de la production et de la solubilisation de protéines recombinantes 6His-PtaGLIM1a dans différentes souches d'E. coli	235
2.2.2.	Purification de la protéine recombinante 6His-PtaGLIM1a sous forme de corps d'inclusion puis par affinité sur une résine de cobalt	237
2.2.3.	Tests de renaturation et de solubilisation de la protéine recombinante 6His-PtaGLIM1a	238
2.3.	<i>Détection avec les anticorps A16 des protéines LIM de peuplier appartenant au groupe LIM1</i>	239
2.4.	<i>Estimation du titre des anticorps anti-6His-GLIM1a et anti-P-GLIM1a sur la protéine recombinante 6His-PtaGLIM1a et sur le peptide P-GLIM1a</i>	241
2.5.	<i>Immunodétection des protéines PtaGLIM1a de peuplier avec les anticorps anti-P-GLIM1a</i>	243
<b>3</b>	<b>Discussion</b>	<b>245</b>
3.1.	<i>Les protéines PtaGLIM1a pourraient former des dimères de protéine in vitro et chez le peuplier</i>	245
3.2.	<i>Les anticorps A16 détectent plusieurs protéines du groupe LIM1 chez le peuplier</i>	247

3.3.	<i>Les anticorps anti-P-GLIM1a confirment l'induction de l'expression du gène PtaGLIM1a dans le bois de tension</i>	249
<b>4</b>	<b>Conclusions</b>	<b>251</b>
<b>Discussion Générale</b>		<b>253</b>
1.	<i>Les protéines LIM de plantes ont une diversité structurale différente de celles des autres eucaryotes</i>	253
2.	<i>Les spécificités d'expression tissulaire des gènes LIM donnent des indications sur leurs fonctions potentielles lors du développement de la plante</i>	254
3.	<i>Les protéines LIM : une double fonction cellulaire ?</i>	258
3.1.	Des protéines se liant à l'actine	258
3.2.	Facteurs de transcription régulant des gènes de la voie de biosynthèse des lignines	263
<b>Conclusions et Perspectives</b>		<b>267</b>
<b>1</b>	<b>Conclusions</b>	<b>267</b>
1.1.	<i>Les gènes LIM de peuplier, d'Arabidopsis et de riz</i>	267
1.2.	<i>Relations entre les spécificités d'expression des gènes LIM de peuplier et la phylogénie de la famille de protéines à domaine LIM chez les plantes</i>	268
1.3.	<i>La protéine PtaGLIM1a est fortement exprimée dans le bois de tension</i>	270
<b>2</b>	<b>Perspectives</b>	<b>271</b>
2.1.	<i>Localisation tissulaire et cellulaire de la protéine PtaGLIM1a</i>	271
2.2.	<i>Détermination de la localisation subcellulaire de la protéine PtaGLIM1a</i>	272
2.3.	<i>La protéine PtaGLIM1a est elle un facteur de transcription et/ou une protéine se liant à l'actine ?</i>	273
2.4.	<i>Fonction de la protéine PtaGLIM1a dans la formation de bois de tension</i>	274
<b>Bibliographie</b>		<b>277</b>
<b>Annexes</b>		<b>297</b>
	<b>Annexe 1.</b> Séquences génomiques corrigées des gènes de Populus trichocarpa codant pour des protéines à domaine LIM	297
	<b>Annexe 2.</b> Informations sur les gènes codant pour des protéines à domaine LIM exprimés chez les plantes	309
	<b>Annexe 3.</b> Classification phylogénétique et taxonomie des espèces de plantes pour lesquelles des gènes LIM ont été identifiées	313

<b>Annexe 4.</b> Arbre phylogénétique de la famille de protéines à domaine LIM chez les plantes (parcimonie)	317
<b>Annexe 5.</b> Arbre phylogénétique de la famille de protéines à domaine LIM chez les plantes (Phyml et BIONJ)	319
<b>Annexe 6.</b> Alignement des séquences protéiques des domaines LIM1 et LIM2	320
<b>Annexe 7.</b> Carte de la séquence du vecteur pENTR™/D-TOPO® utilisé pour le clonage de la séquence codante du gène PtaGLIM1a	323
<b>Annexe 8.</b> Carte de la séquence des vecteurs de destination pDEST™15 et pDEST™17 respectivement utilisé pour la production des protéine recombinantes GST-PtaGLIM1a et 6His-PtaGLIM1a	323
<b>Annexe 9.</b> Article publié dans <i>DNA Research</i>	325



## Liste des abréviations

ABP : Actin binding protein	PGPS/D1 : petunia germinating pollen clone S/D1
ADN : Acide désoxyribonucléique	PHYLIP : Phylogenetic Interference Package
ADNc : ADN complémentaire	pI : point isoélectrique
AFLP : Amplified Fragment Length Polymorphism	PKC : Protéine Kinase C
AGP : Arabinogalactan protein	PKL : pickle
APS : ammonium persulfate	PLIM : Pollen specific LIM domain protein
ARN : acide ribonucléique	PVP : polyvinyl pyrrolidone
ARNm : ARN messenger	RT : Reverse Transcription
ASN : asparagine	S : unité syringyle
ATP : adénosine triphosphate	SF3 : sunflower specific 3
BLAST : Basic Local Alignment Search Tool	SDS : Sodium Dodécyl Sulfate
BO : Bois opposé	SNP : Single Nucleotide Polymorphism
BSA : Bovin serum albumin	SuSy : sucrose synthase
BT : Bois tendu	TAIR : The Arabidopsis Information Resource
BY2 : Bright Yellow 2	TCA : acide trichloroacétique
CAD : cinnamyl alcool déshydrogénase	TEMED : Tétraméthyl-éthylènediamine
CDS : Coding Sequence	TIGR : The Institute for Genomic Research
CesA : Cellulose synthase	TRIS : Tris hydroxyméthyl aminométhane
CKII : Caséine kinase 2	UBC : University of British Columbia
CNRS : Centre National de la Recherche Scientifique	UTR : Untranslated Region
CR : Coding Region	XD : xylème en développement
CRP : Cystein Rich Protein	XM : xylème mature
CTAB : bromure d'hexadécyltriméthylammonium	WAK : Wall Associated Kinase
DEPC : diéthyl pyrocarbonate	WLIM : Widely expressed LIM protein
DTT : dithiothréitol	ZC : Zone cambiale
DO : densité optique	4CL : 4-coumarate:CoA ligase
dNTP : déoxynucléotide triphosphate	6His : 6 Histidines
EBI : European Bioinformatics Institute	
EDTA : acide éthylènediaminetétraacétique	
EMBL : European Molecular Biology Laboratory	
EST : Expressed Sequence Tag	
FLA : Fasciclin-like arabinogalactan protein	
FLIM : Fibre specific LIM domain protein	
Fibre G : fibre gélatineuse	
G : unité guaïacyle	
GFP : Green Fluorescent Protein	
GRR : Glycine-Rich Region	
GST : Glutathione S Transférase	
HD : Homéodomaine	
INRA : Institut National de la Recherche Agronomique	
IRD : Institut de Recherche pour le Développement	
ISL-1 : Insulin gene enhancer binding protein	
JTT : matrice Jones Taylor Thornton	
LIM : <u>L</u> IN-11, <u>I</u> SL-1, <u>M</u> EC-3	
LIN-11 : <i>Caenorhabditis elegans</i> cell lineage gene	
MEC-3 : LIM-homeobox-containing gene <i>MEC-3</i>	
MLP : Muscle LIM Protein	
MM : Masse Moléculaire	
MS : milieu Murashige & Skoog	
MYB : facteur de transcription homologue aux proto-oncogène <i>c-MYB</i>	
ORF: Open Reading Frame	
NCBI : National Center for Biotechnology Information	
PAL : phénylalanine ammonia lyase	
PCR : Polymerase Chain Reaction	



## Introduction générale

### 1 Le peuplier

Le nom peuplier vient de *Populus* qui signifie peuple en latin. En effet, les peupliers étaient plantés par les Romains sur les lieux publics. Traditionnellement, les peupliers ont été plantés comme brise-vent ou pour contrôler l'érosion des sols. Bien que le peuplier soit économiquement moins important que les espèces de pin ou l'eucalyptus, les espèces du genre *Populus* sont maintenant l'un des groupes d'arbres le plus exploité commercialement et le plus planté à travers le monde. Le genre *Populus* englobe environ 30 espèces largement distribuées dans divers habitats des régions sub-tropicales, tempérées et froides de l'hémisphère nord (Eckenwalder, 1996). En Amérique du Nord, les peupliers s'étendent du Mexique jusqu'à l'Alaska, et en Eurasie, ils se retrouvent de l'Europe occidentale jusqu'à la Chine. Aucun peuplier ne pousse à l'état sauvage dans l'hémisphère Sud, mais il est cultivé par l'Homme dans les régions tempérées de l'Argentine ou du Chili. Le peuplier comprend aussi de nombreux hybrides naturels ou créés par l'homme. En dehors des espèces spontanées, de nombreux cultivars sont à la disposition des populiculteurs. Les peupliers, arbres à croissance rapide, se rencontrent rarement en forêt traditionnelle dans nos régions. Beaucoup d'espèces croissent sur les terrains humides voire temporairement inondés dans les vallées et plaines alluviales. D'autres espèces, comme les peupliers trembles, colonisent des terres élevées en altitude.

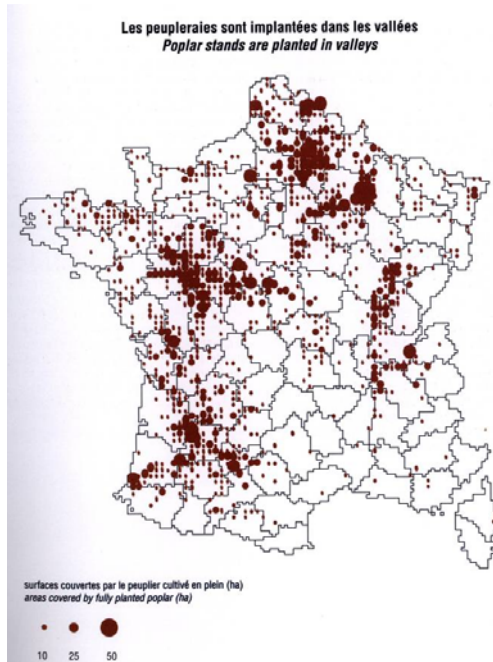
#### 1.1. Importance économique en France

La populiculture est à la frange de l'agriculture et de la sylviculture. En France, les peupleraies cultivées représentent environ 230000 hectares et assurent de 2002 à 2004 une production annuelle d'environ 1,4 millions de m<sup>3</sup> de bois d'œuvre (Agreste, 2004). C'est, avec le hêtre, la deuxième essence feuillue après le chêne en terme de production, et la France est le premier producteur européen de peuplier. La plupart des cultivars ou clones plantés actuellement en France proviennent de pays voisins (Italie, Pays-Bas, Belgique). Ces cultivars appartiennent pour l'essentiel aux espèces *P. deltoides*, *P. trichocarpa*, et *P. nigra*. Les principales zones de production de peupliers se trouvent dans les vallées telles que celles de la Loire, la Garonne, la Saône ou l'Oise, où l'on rencontre des milieux alluviaux répondant à



## Introduction Générale

l'exigence en eau élevée du peuplier (Fig. 1). Les régions ayant le plus de surface cultivées sont les régions Poitou-Charentes, Centre, Picardie, et Aquitaine. Dans les Pays de la Loire, le peuplier occupe une place importante et est la première essence en terme de production de bois d'œuvre avec une récolte de 155800 m<sup>3</sup> en 2005 (Ministère de l'Agriculture et de la Pêche, 2006). Pour la production de bois, la révolution du peuplier est généralement de 20 ans entre la plantation et l'abattage final. La récolte du peuplier est partiellement exportée et sa transformation alimente une part importante de l'industrie du déroulage. Le bois de peuplier est un matériau polyvalent, ce qui autorise son emploi dans des débouchés industriels très diversifiés tels que la filière du déroulage (panneaux contreplaqués, emballages légers, allumettes) et la filière du sciage (meubles, tablettes et rayonnages, caisserie, charpentes légères, paletterie, literie). Avec le développement de la filière bois énergie, les organismes de recherche, notamment l'INRA, se sont intéressés à la conduite du peuplier en taillis à courte rotation (6 à 8 ans) et à très courte rotation (2 à 3 ans). Récoltée grâce à une ensileuse adaptée, la peupleraie est broyée finement pour sa production de biomasse. Abandonné du fait d'une faible rentabilité, la flambée du coût de l'énergie devrait relancer les essais sur ce mode de conduite qui n'est d'ailleurs pas spécifique au peuplier.



**Figure 1.** Implantation des peupleraies en France dans les vallées fluviales en hectare de surfaces couvertes (Ministère de l'Agriculture et de la Pêche, 2000).

### 1.2. Les peupliers appartiennent à la famille des *Salicaceae* et au genre *Populus*

Les peupliers sont des plantes supérieures ligneuses angiospermes, appartenant à la classe du noyau des vraies dicotylédones (core eudicotyledons), à la sous-classe des rosidées et plus précisément à la clade des eurosids I. En comparaison, *Arabidopsis thaliana* de la famille des *Brassicaceae* est incluse dans la clade des eurosids II. Selon l'ancienne classification taxonomique, l'ordre des *Salicales* ne comprenait qu'une seule famille de plantes dicotylédones : les *Salicaceae* dont les deux genres les plus importants sont les saules (*Salix*) et les peupliers (*Populus*). Dans une analyse cladistique plus récente des angiospermes, cet ordre des *Salicales* n'existe plus et a été remplacé par l'ordre des *Malpighiales* (The Angiosperm Phylogeny Group, 2003). Cet ordre des *Malpighiales* comprend au moins 29 familles dont les familles des *Salicaceae* et des *Flacourtiaceae* très proches entre elles, et la famille des *Euphorbiaceae* comprenant le manioc (*Manihot esculenta*) et les euphorbes (*Euphorbia* sp). L'ancêtre commun des deux genres *Salix* et *Populus* serait apparu il y a 65 millions d'années (Eckenwalder, 1996). Les peuplier tremble, baumier et deltoïdes sont apparus plus tard à la fin du Miocène, il y a 5-10 millions d'années lors d'un refroidissement global de la terre. Les arbres de la famille des *Salicaceae* ont des caractéristiques communes comme par exemple, une croissance rapide, des feuilles simples, alternes, caduques, des fleurs mâles ou femelles regroupées en chatons et de nombreuses graines de petite taille entourées de poils ou coton. Parmi les arbres tempérés avec des chatons femelles, seuls les peupliers et les saules ont de telles graines avec des fibres cotonneuses (Eckenwalder, 1996). Ce sont des espèces généralement dioïques, c'est à dire avec des individus soit mâles soit femelles. Les deux genres *Populus* et *Salix* diffèrent notamment par le port du chaton, dressé (au moins avant l'anthèse) chez les saules et pendant chez les peupliers, par le nombre d'étamines, élevé chez les peupliers et faible chez les saules. Les feuilles de peuplier sont généralement larges avec des nervures palmatopennées, et celles des saules sont étroites et strictement pennées. Les deux genres *Salix* et *Populus* sont bien séparés entre eux et le genre *Populus* est monophylétique, tandis que le genre *Salix* serait plutôt paraphylétique (Leskinen & Alström-Rapaport, 1999).

**Tableau 1.** Classification taxonomique des espèces du genre *Populus* (Eckenwalder, 1996)

Section	Espèce	Origine géographique
<i>Abaso</i>	<i>P. mexicana</i> Wesmael	Amérique du Nord
<i>Turanga</i> Bge.	<i>P. euphratica</i> Olivier <i>P. ilicifolia</i> (Engler) Rouleau <i>P. pruinosa</i> Schrenck	Eurasie, Afrique du Nord Afrique de l'Est Asie
<i>Leucoides</i> Spach.	<i>P. lasiocarpa</i> Olivier <i>P. glauca</i> Haines <i>P. heterophylla</i> L.	Asie Asie Amérique du Nord
<i>Aigeiros</i> Duby	<i>P. fremontii</i> S. Watson <i>P. deltoides</i> Bartr. ex Marsh. <i>P. nigra</i> L.	Amérique du Nord Amérique du Nord Eurasie, Afrique du Nord
<i>Tacamahaca</i> Spach.	<i>P. augustifolia</i> James <i>P. balsamifera</i> L. <i>P. ciliata</i> Royle <i>P. laurifolia</i> Ledebour <i>P. simonii</i> Carrière <i>P. suaveolens</i> Fischer <i>P. szechuanica</i> Schneider <i>P. trichocarpa</i> Torr. & A. Gray <i>P. yunnanensis</i> Dode	Amérique du Nord Amérique du Nord Asie Eurasie, Afrique du Nord Asie Asie Asie Amérique du Nord Asie
<i>Populus/Leuce</i> Duby	<i>P. adenopoda</i> Maximowicz <i>P. alba</i> L. <i>P. gamblei</i> Haines <i>P. grandidentata</i> Michaux <i>P. guzmanantlensis</i> Vazquez & Cuevas <i>P. monticola</i> Brandegee <i>P. sieboldii</i> Miquel <i>P. simaroa</i> Rzedowski <i>P. tremula</i> L. <i>P. tremuloides</i> Michaux	Asie Eurasie, Afrique du Nord Asie Amérique du Nord Amérique du Nord Amérique du Nord Amérique du Nord Asie Amérique du Nord Eurasie, Afrique du Nord Amérique du Nord

Le nombre d'espèces de peuplier (*Populus* spp.) jusqu'alors décrit dans la littérature varie de 22 à 85, avec plus d'une centaine d'hybrides et de cultivars (Eckenwalder, 1996). Les désaccords dans le nombre d'espèces peuvent être liés à la mésinterprétation de certains hybrides et à la difficulté de délimiter les frontières entre espèces. Actuellement, 29 espèces du genre *Populus* ont été divisées en six sections botaniques (Tableau 1) selon des critères morphologiques, écologiques et d'hybridations interspécifiques (Eckenwalder, 1996). Les variations dans les caractéristiques anatomiques entre espèces de peuplier marquent les différences taxonomiques au sein de ce genre, et les variations de la forme ovale des feuilles de peuplier sont des différences prédominantes. De plus, les espèces peuvent être croisées entre elles au sein d'une même section. Par exemple, pour la section *Populus*, des hybrides tels que *P. tremula* × *P. tremuloides* ou *P. tremula* × *P. alba* ont été sélectionnés. Les études phylogénétiques des espèces du genre *Populus* sont en général consistantes avec la classification taxonomique établie par Eckenwalder en 1996 (Hamzeh & Dayanandan, 2004 ; Cervera *et al.*, 2005). Toutefois des exceptions ont été observées, comme par exemple l'espèce *P. nigra* de la section *Aigeiros* qui pourrait être un ancien hybride naturel. En effet,

## Introduction Générale

une étude phylogénétique récente placerait cette espèce dans la section *Leuce/Populus* selon l'analyse de l'ADN chloroplastique et dans la section *Aigeiros* selon l'analyse de l'ADN nucléaire (Hamzeh & Dayanandan, 2004). Suite à des analyses phylogénétiques avec des marqueurs AFLP nucléaires, d'autres auteurs proposent pour *P. nigra* la création d'une nouvelle section *Nigrae* séparée de l'espèce *P. deltoides* (Cervera *et al.*, 2005). La relation entre les sections *Aigeiros* et *Tacamahaca* est particulière. Bien que provenant de niches écologiques différentes, les espèces de ces deux sections se ressemblent fortement au niveau des fleurs et peuvent être croisées entre elles. Ainsi, les *P. deltoides* peuvent être croisés avec les *P. nigra*, *P. balsamifera* et *P. trichocarpa* pour donner des hybrides fertiles (Eckenwalder, 1996). De plus, de fortes relations phylogénétiques ont été observées entre les espèces de ces deux sections (Hamzeh & Dayanandan, 2004 ; Cervera *et al.*, 2005). Contrairement aux autres sections (*Abaso*, *Turanga*, *Leucoides* et *Aigeiros*), les sections *Tacamahaca* et *Populus/Leuce* comprennent le plus d'espèces et sont très étendues sur les continents. Les peupliers baumiers (*P. trichocarpa*) sont plus spécifiquement implantés en Amérique du Nord et les peupliers trembles (*P. tremula*) sont répartis en Eurasie et Afrique du Nord (Tableau 1). Les sections *Leuce/Populus*, *Tacamahaca* et *Aigeiros* comprennent des espèces d'importance économique. En effet, les peupliers deltoïde (*P. deltoides*), baumier et noir (*P. nigra*) sont les espèces de peuplier les plus exploitées dans le monde. Parmi les espèces du genre *Populus*, seules trois sont considérées comme indigènes en France : le peuplier tremble, le peuplier blanc (*P. alba*) et le peuplier noir.

### 1.3. Développement et morphologie du peuplier

Les espèces du genre peuplier sont toutes des arbres à feuilles caduques. Les peupliers font partie des arbres tempérés ayant la croissance la plus rapide. Certains arbres peuvent faire 40 m de hauteur en moins de 20 années (Bradshaw *et al.*, 2000). Tout le long de la saison de croissance, de février/mars à octobre, les tiges de peuplier continuent de pousser en initialisant de nouvelles feuilles, même après l'éclosion des premiers bourgeons de feuilles préformées avant l'hiver. Les tiges initient séquentiellement les bourgeons végétatifs précoces, les bourgeons floraux et les bourgeons végétatifs tardifs (Yuceer *et al.*, 2003). Les feuilles sont alternes, stipulées, pétiolées et simples avec des dents glandulaires le long des bords et souvent des glandes à la jonction du limbe et du pétiole qui a un aspect aplati (Eckenwalder, 1996). Le bois de peuplier est poreux, diffus, clair et léger. C'est en partie la conséquence

## Introduction Générale

d'une croissance rapide, presque continue pendant les saisons favorables (printemps, été). La formation du bois de peuplier sera expliquée plus tard dans le détail dans un chapitre spécifiquement dédié au processus complexe de la xylogénèse. Comparé aux autres arbres, l'écorce des peupliers reste fine pendant une longue période, surtout chez les peupliers trembles. Toutes les espèces de peuplier ont la capacité de se reproduire de manière asexuée ou clonale, la plupart du temps par bourgeonnement au niveau du collet de la racine d'arbres morts, ou par enracinement de tiges (Eckenwalder, 1996). Leur système racinaire est important et invasif. La plupart des peupliers peuvent aussi se propager par des drageons de tiges portés par les racines horizontales.

Jusqu'à présent, les organes reproducteurs mâles ou femelles de peuplier ont été très peu étudiés au niveau moléculaire. Seul, le développement et la morphologie des fleurs ont été décrits dans le détail (Boes & Strauss, 1994 ; Kaul, 1995 ; Yuceer *et al.*, 2003). Les peupliers sont des espèces dioïques avec des arbres mâles ou femelles à l'exception de *P. lasiocarpa* (Boes & Strauss, 1994). Des fleurs bisexuées ont également été observées de façon occasionnelle chez d'autres espèces de peuplier comme par exemple *P. tremula* ou *P. trichocarpa*. Les fleurs de peuplier ne portent pas de pétales ni de sépales et comportent uniquement deux pièces florales : une coupe périanthaire réduite et des étamines ou des carpelles. D'autre part, les fleurs mâles et femelles sont bien vascularisées par des fibres vasculaires discrètes au niveau des parties périanthaires (Kaul, 1995). Les méristèmes d'inflorescences situés à la jonction des pétioles et des tiges, initient leur développement précocément au cours du printemps. Chez *P. trichocarpa*, l'initiation débute trois à quatre semaines après l'anthèse des fleurs de l'année en cours (Boes & Strauss, 1994). Toutefois, les fleurs mâles ou femelles ne peuvent se distinguer anatomiquement qu'à la fin du printemps. Leur développement se poursuit durant l'été jusqu'à l'automne pendant la deuxième phase de croissance de l'arbre, puis ces bourgeons entrent en dormance pendant l'hiver pour éclore l'année suivante (Boes & Strauss, 1994 ; Kaul, 1995). La mégasporogénèse et la microsporogénèse débutent tardivement au cours de l'hiver, environ deux semaines avant l'anthèse (Boes & Strauss, 1994). A partir de ces bourgeons spécialisés contenant les inflorescences préformées, les peupliers fleurissent au début du printemps, en général avant l'émergence des feuilles précoces. Dans la plupart des régions, la date de floraison varie entre espèces de peuplier vivant dans le même milieu. Par exemple, dans les pépinières de l'INRA d'Orléans, les clones femelles de *P. trichocarpa* et de *P. nigra* fleurissent avant les clones femelles de *P. deltoides*. Le moment de l'anthèse varie aussi entre différents arbres de la même espèce (Boes & Strauss, 1994).



**Figure 2.** Morphologie des fleurs ou chatons mâles après anthèse

Inflorescences mâles provenant de différentes espèces de peupliers : *Populus deltoides* (A), *P. nigra* (B), et le cultivar Dorskamp (*P. deltoides* × *P. nigra*) (C). Les chatons mâles sont pendants et portent des étamines rouges à différents stades de développement. Les branches portant ces inflorescences mâles ont été prélevées sur des peupliers cultivés ou sauvages poussant en bord de Loire.

Étamines de *P. nigra* à différents stades de développement et observées à la loupe binoculaire : précoce (D et F), tardif (E et G), et mature (H) avec des grains de pollen de couleur jaune. La flèche en E montre le périanthe portant les étamines. F et G illustrent des étamines composées de quatre anthères.

### 1.3.1. Morphologie et développement des inflorescences mâles

Les organes reproducteurs mâles sont composés de nombreuses étamines portant un filament court peu visible, et quatre anthères de couleur rouge foncé (Fig. 2F et 2G). Les étamines sont supportées par un périanthe de couleur jaune en forme de disque (Fig. 2E). Le nombre d'étamines par chaton varie de 5 à 60 entre les espèces de peuplier (Eckenwalder, 1996). La morphologie et la couleur des chatons mâles varient nettement entre les espèces de peuplier. Ainsi, chez le cultivar Dorskamp (*P. deltoides* × *P. nigra*) les chatons sont beaucoup plus allongés et présentent une couleur rouge plus foncée par rapport à ceux des espèces *P. deltoides* et *P. nigra* (Fig. 2A à C). Le développement des étamines commence au centre

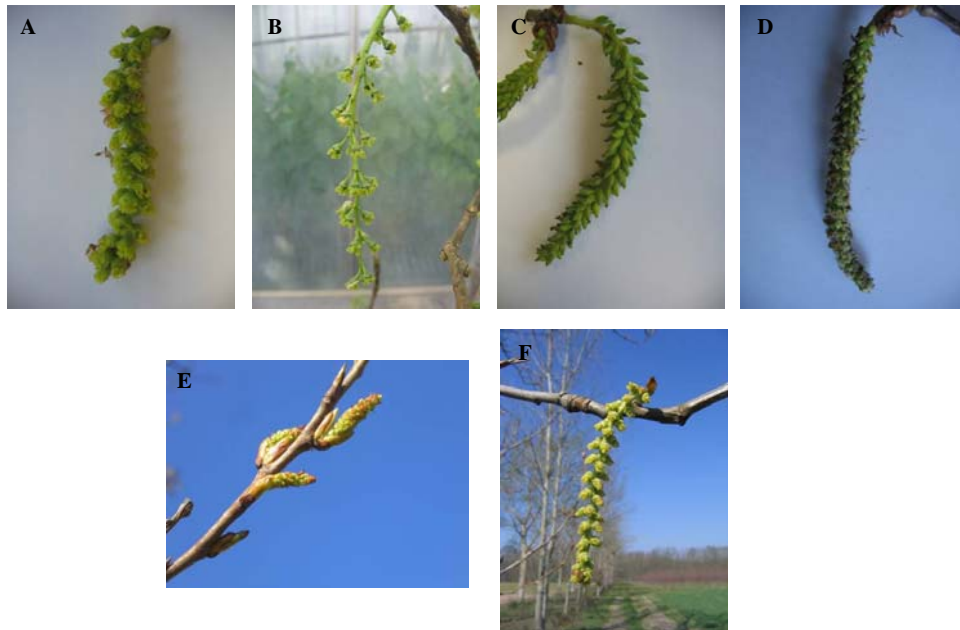
d'une région méristématique en forme de disque et continue de manière centrifuge (Boes & Strauss, 1994). A l'anthèse, les écailles des bourgeons floraux s'ouvrent, et les étamines en forme de grappe sont disposées de façon sessile et entassées sur le rachis (Fig. 2D). Ensuite, au cours du développement des étamines, la taille des chatons mâles augmente et leur port devient pendant (Fig. 2E). Le rachis et les pédoncules s'allongent considérablement durant cette phase, les étamines sont plus étalées, et les écailles des bourgeons tombent. A la fin de leur développement, les grains de pollen se déshydratent, les étamines s'ouvrent puis deviennent déhiscentes pour faciliter la libération du pollen mature (Fig. 2G et H). La dispersion du pollen et la pollinisation s'effectuent par le vent. Les peupliers sont donc anémophiles contrairement aux espèces du genre *Salix* (Eckenwalder, 1996).

### 1.3.2. Morphologie et développement des inflorescences femelles

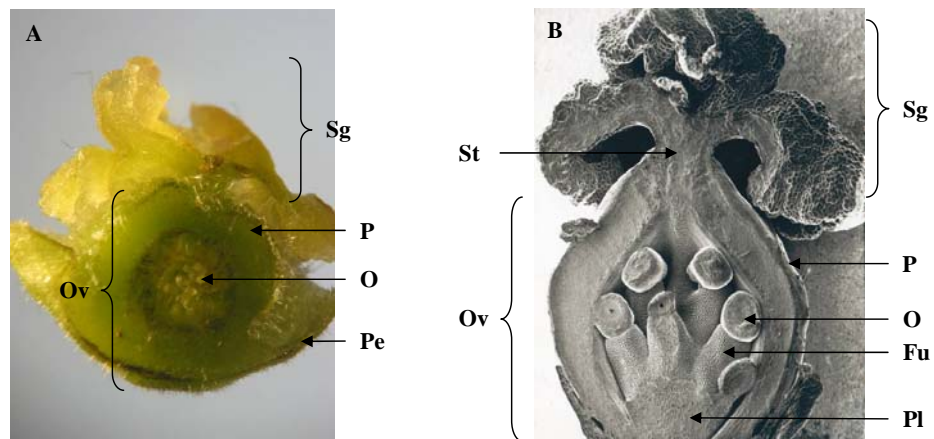
Les organes reproducteurs femelles contiennent un ovaire uniloculaire avec un nombre de carpelles variable (de 2 à 4) entre les sections du genre *Populus* (Eckenwalder, 1996). A l'anthèse, l'inflorescence femelle est une grappe de fleurs entassées plus ou moins érigée (Fig. 3E). Comme pour les fleurs mâles, après éclosion des bourgeons femelles, ceux-ci se développent et augmentent de taille en s'allongeant par le rachis, puis les fleurs femelles deviennent pendantes (Fig. 3F). Chaque fleur est sous-tendue par une bractée d'aspect frimbrié comme celles des fleurs mâles (Fig. 3E), mais chez les fleurs femelles les pédoncules s'allongent peu, les bractées restent sessiles sur le rachis, et elles ne s'élèvent pas avec la maturation de la fleur (Kaul, 1995). Lorsque les inflorescences émergent et se développent, les carpelles se découvrent en poussant de côté les bractées qui finiront par tomber. La fleur femelle mature est alors composée d'un ovaire globulaire avec un péricarpe vert foncé couvert de trichomes, et surmonté par des stigmates vert clair profondément lobés portés par un style très court encore appelé stylode (Fig. 4A et 4B). Chez *P. nigra* le stigmate est composé de quatre lobes stigmatiques. L'ovaire est partiellement recouvert d'un périanthe vert foncé formant une crête autour de lui. Il contient de nombreux ovules anatropes attachés au placenta par un funicule épais (Fig. 3B) (Villar *et al.*, 1993 ; Boes & Strauss, 1994). Les fleurs femelles ont des phénotypes différents selon les espèces de peuplier. Chez *P. trichocarpa*, les ovaires et stigmates sont plus gros comparés à *P. deltoides* et *P. alba* (Fig. 3A et 3B). Les stigmates sont très petits et l'ovaire a une forme ovoïde dans le cas de l'espèce *P. alba* (Fig.

## Introduction Générale

3C), et chez le clone hybride 717-1-B4 (*P. tremula* × *P. alba*), les fleurs femelles sont petites mais nombreuses (Fig. 3D).



**Figure 3.** Morphologie des fleurs ou chatons femelles après l'anthèse  
 Inflorescences femelles prélevées sur différentes espèces de peupliers : *Populus trichocarpa* (A), *P. deltoides* (B), *P. alba* (C), et le clone hybride 717-1-B4 (*P. tremula* × *P. alba*) (D). La morphologie des fleurs femelles varie considérablement entre les différentes espèces. Les branches portant les inflorescences femelles ont été prélevées sur des peupliers cultivés en pépinière à l'INRA d'Orléans.  
 Fleurs femelles à deux stades de développement différents : juvénile chez *P. nigra* (E) et mature chez *P. trichocarpa* pouvant être fécondé (F). Les chatons femelles d'abord érigés deviennent pendants au cours de leur développement.

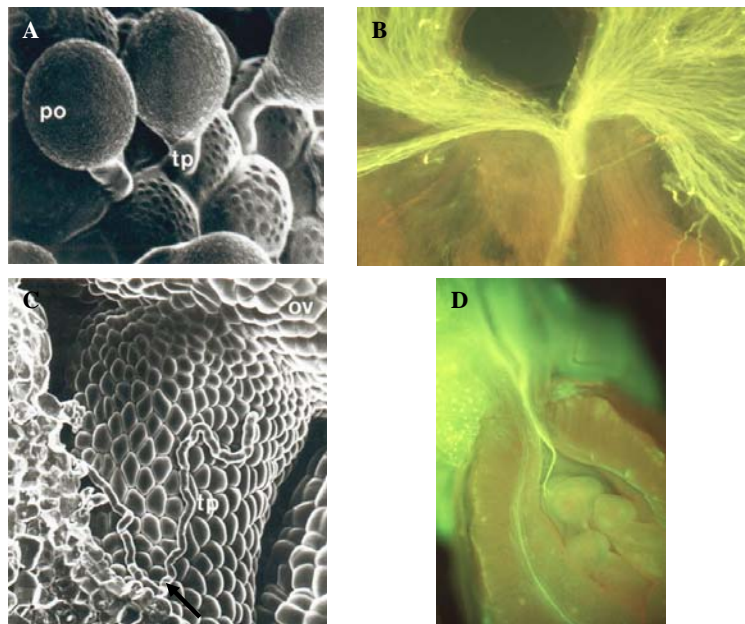


**Figure 4.** Détails des structures reproductives au sein d'une fleur femelle de peuplier  
 Fleurs femelles de peuplier coupées longitudinalement et observées à la loupe binoculaire (*Populus trichocarpa*) (A), et en microscopie électronique à balayage (*P. nigra*) (B) (Villar et al. 1993). Le stigmate (Sg) qui réceptionne le pollen est porté par le style (St) situé au dessus de l'ovaire (Ov). Ce dernier est protégé par le péricarpe (P) et contient plusieurs ovules (O). L'ovule est relié au placenta (Pl) par le funicule (Fu), et à la base de l'ovaire, se trouve le périanthe (Pe).



### 1.3.3. Fécondation et formation du tube pollinique

Lors de la fécondation, les grains de pollen déposés sur la surface des stigmates par le vent, s'hydratent et germent en formant un tube pollinique (Fig. 5A). La formation du tube pollinique est un processus de croissance cellulaire unique aux plantes supérieures. Elle est caractérisée par une expansion cellulaire particulière où l'élongation cellulaire anisotrope est localisée à l'extrémité du tube pollinique. Cette élongation intrusive du tube pollinique permet la pénétration des tissus du stigmate, et la traversée du style jusque dans la cavité ovarienne pour délivrer les gamètes mâles aux ovules (Fig. 5A à 5D). Chez *Populus nigra*, le processus d'élongation du tube pollinique depuis la pollinisation sur le stigmate jusqu'à la rencontre avec le micropyle de l'ovule dure environ 25 h (Villar *et al.*, 1987). Il est remarquable d'observer les tubes polliniques se joindre ensemble au niveau du style (Fig. 5B), leur croissance le long de la paroi interne de la cavité ovarienne, et leurs changements de direction vers l'ovule le long du funicule (Fig. 5C et 5D). Chez le peuplier, aucune étude n'a encore été réalisée sur les gènes impliqués dans le développement du pollen et du tube pollinique.



**Figure 5.** Fécondation des ovules par le pollen et formation du tube pollinique chez *Populus nigra* (Villar *et al.* 1987)

(A) Détails de grains de pollen germent sur la surface stigmatique. Les grains de pollen germent en formant un tube pollinique qui pénètre dans le stigmate. (B) Faisceaux de tubes polliniques dans les lobes stigmatiques se rejoignant au niveau du style à la base de la figure. (C) Tubes polliniques en croissance le long du funicule de l'ovule. A la base du funicule, le tube pollinique change de direction (marqué par une flèche) pour aller vers l'ovule. (D) Faisceaux de tubes polliniques atteignant la cavité ovarienne avec leurs extrémités au niveau des ovules. La progression du tube pollinique a été observée en microscopie électronique à balayage (A et C) et en microscopie à fluorescence après coloration au bleu d'Aniline (B et D). Grains de pollen (po), tube pollinique (tp), ovule (ov).

## Introduction Générale

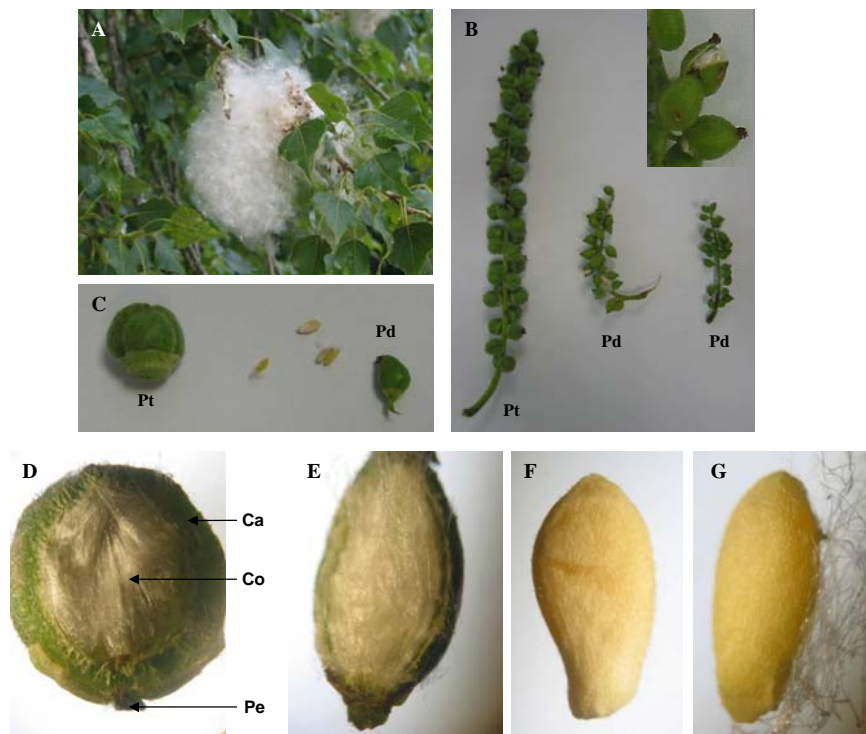
La formation du tube pollinique implique des mécanismes moléculaires particuliers pour le transport des métabolites nécessaires à l'élaboration de la paroi cellulaire à l'extrémité du tube. Le cytosquelette d'actine et dans une moindre mesure celui de microtubule sont essentiels à la germination et à la croissance polarisée du tube pollinique (Taylor & Hepler, 1997). En effet, les microfilaments d'actine, agencés parallèlement à l'axe du tube pollinique, s'accumulent dans la zone sub-apicale du tube pour former une matrice dense, et interviendraient avec la myosine dans le transport de vésicules contenant les éléments essentiels à l'élaboration de la paroi cellulaire (Taylor & Hepler, 1997 ; Vidali & Hepler, 2001). De plus, des expériences utilisant des agents favorisant la dépolymérisation de l'actine (profilin, latrunculin B et cytochalasin D) à des conditions permettant toujours le transport des vésicules cytoplasmiques, montrent que la polymérisation des monomères d'actine en filaments est nécessaire à la croissance du tube pollinique (Gibbon *et al.*, 1999 ; Vidali *et al.*, 2001 ; Justus *et al.*, 2004 ; Gossot & Geitmann, 2007). Les microtubules, organisés parallèlement à l'axe du tube pollinique, sont également importants dans le transport des organelles le long du tube pollinique, notamment par l'intermédiaire de protéines motrices apparentées aux kinésines (Romagnoli *et al.*, 2003). Alors que chez les angiospermes les agents inhibiteurs de microtubules ont peu d'effet sur la croissance du tube pollinique, chez les gymnospermes, ces agents bloquent partiellement la germination et la croissance générant entre autre des tubes polliniques plus courts (Anderhag *et al.*, 2000 ; Hepler *et al.*, 2001). D'autre part, pour certaines espèces comme le pétunia ou le maïs, les flavonols apparaissent nécessaires à la germination du tube pollinique. Il a été montré chez des mutants de pétunia et de maïs déficients en chalcone synthase et ne synthétisant plus de flavonoïdes que la germination du pollen et la fertilisation sont fortement perturbés (Mo *et al.*, 1992). Toutefois, les flavonols sont beaucoup moins nécessaires à la germination du pollen chez *Arabidopsis thaliana* (Taylor & Hepler, 1997). Enfin, les protéines à arabinogalactane (AGP), synthétisées par le grain de pollen ou le pistil, jouent également un rôle important dans le développement du tube pollinique (Cheung *et al.*, 1995 ; Roy *et al.*, 1998).

### 1.3.4. Morphologie et développement des organes fructifères

Après fécondation, les fleurs femelles donnent des capsules vertes à paroi fine et contenant les graines entourées de poils cotonneux (Fig. 6A à 6C). Ces fibres de coton aident à la dissémination des graines par le vent. Les fruits s'agrandissent rapidement après la

## Introduction Générale

pollinisation et les stigmates flétrissent. Les poils cotonneux unicellulaires ne se développent qu'après la fécondation et proviennent du placenta pariétal (Boes & Strauss, 1994 ; Kaul, 1995). Le nombre de graines par capsule (de 2 à 40) varie entre les sections de peuplier (Eckenwalder, 1996). Les inflorescences, capsules et graines présentent différentes formes en fonction des espèces de peuplier. Les inflorescences des peupliers *P. trichocarpa* sont en général beaucoup plus développées que celles de *P. deltoides* (Fig. 6B). Chez *P. trichocarpa*, les capsules d'aspect très arrondi sont beaucoup plus grosses que chez *P. deltoides* d'aspect plus ovoïde (Fig. 6C à 6E), et les graines sont plus aplaties chez *P. trichocarpa* (Fig. 6F et 6G). Les fruits s'assèchent à la fin du printemps, deviennent déhiscents et perdent leur couleur verte. Après avoir libéré le pollen ou les graines, les inflorescences tombent et sont perdues (Kaul, 1995). Les graines de peuplier ne sont ni dormantes, ni résistantes, et une fois tombée au sol, elles doivent germer immédiatement. Un sol ayant une forte humidité est nécessaire à la germination de ces graines. Typiquement, les graines de peupliers noirs poussant en bordure de la Loire ne germeront que si les rives du fleuve ont été inondées par une crue de printemps (Villar M., communication personnelle).



**Figure 6.** Morphologie des organes fructifères de peuplier (A) Peuplier (*P. trichocarpa*) portant une fructification de graines entourées de poils cotonneux (Photo M. Villar). (B) Différents organes fructifères de peupliers *P. trichocarpa* (Pt) et *P. deltoides* (Pd) avec en encadré le grossissement d'une capsule s'ouvrant pour libérer les gaines entourées de coton. (C) Détails des capsules et graines de *P. trichocarpa* et *P. deltoides*. Capsules (D et E) et graines (F et G) de *P. trichocarpa* (D et F) et *P. deltoides* (E et G) observées à la loupe binoculaire. La capsule (Ca) issue du péricarpe de l'ovaire et portée par le pédoncule (Pe), contient de nombreux poils cotonneux (Co) masquant les graines.

### 1.4. Pourquoi avoir choisi le peuplier (*Populus trichocarpa*) comme arbre modèle ?

La forme de vie des arbres impose des adaptations physiologiques et morphologiques différentes comparées à celles des plantes annuelles. En effet, contrairement à celles-ci, les arbres produisent une grande quantité de xylème secondaire encore appelé bois. De plus, la phénologie de la croissance et de la floraison des arbres est particulière, et ceux-ci présentent deux stades de développement juvénile et mature au cours de leur vie. De même, la pérennité des arbres, leur grande taille impliquant un transport d'eau et de nutriments sur de longues distances, et leur nécessité de vivre longtemps dans un milieu fixe soumis à des changements biotiques et abiotiques continuels sont d'autres caractéristiques écophysiologicals spécifiques aux arbres, et ne pouvant être étudiées facilement chez d'autres plantes modèles comme *Arabidopsis thaliana* et le riz (*Oryza sativa*). C'est pourquoi la nécessité d'un arbre modèle pour la génomique fonctionnelle est devenue une évidence. Le genre *Populus* présente beaucoup de critères expliquant son choix comme arbre modèle :

- Tout d'abord, il est important de préciser que les arbres ne forment pas un groupe monophylétique au sein des plantes supérieures. En effet, ils sont retrouvés dans différents genres et familles de plantes, et sont apparus de multiples fois au cours de l'évolution (Jansson & Douglas, 2007). Le peuplier est plus proche phylogénétiquement d'*Arabidopsis* que ne le sont d'autres arbres comme les pins (des gymnospermes) ou même l'eucalyptus faisant partie de l'ordre des *Myrtales* de la sous-classe des rosidées. Les genres *Populus* et *Arabidopsis* appartiennent aux sous-classes des eurosidées I et II respectivement. Les études de biologie comparée entre ces deux espèces modèles représentant deux genres de dicotylédones extrêmement différentes, permettront l'identification de mécanismes conservés entre ces deux plantes. Mais surtout elles permettront d'analyser et de comprendre les mécanismes moléculaires impliqués dans des processus physiologiques spécifiques aux arbres comme la formation du bois par exemple. De plus, ces études de biologie comparée révéleront les différences génétiques expliquant comment deux lignées soumises à différentes pressions de sélection ont utilisé des mécanismes cellulaires communs de différentes manières. Enfin, la comparaison entre les génomes de peuplier et d'*Arabidopsis* pourrait nous éclairer sur les caractéristiques de leurs ancêtre commun, l'eurosidé ancestral vivant il y a 100-120 millions d'années (Wikström *et al.*, 2001 ; Sanderson *et al.*, 2004).

- D'autre part, il est facile de propager sexuellement les espèces du genre *Populus* par des hybridations intraspécifiques et aussi interspécifiques. Le pollen peut être conservé

## Introduction Générale

pendant plusieurs années et être utilisé pour féconder des branches excisées portant des chatons femelles. Des centaines de graines peuvent être produites en 4-8 semaines. La germination des graines ne dure que 24 heures et elles peuvent fournir des arbres de 1 à 2 m au cours de l'année (Bradshaw *et al.*, 2000). Les peupliers présentent une hérédité diploïde ( $2n = 38$ ), et les espèces du genre peuvent être hybridées entre elles, même au sein de sections différentes. Les hybrides F1 présentent souvent une vigueur hybride ou hétérosis avantageuse commercialement. En revanche, contrairement à *Arabidopsis thaliana*, une plante autogame, la création de lignées pures ou consanguines est difficile chez le peuplier. En effet les peupliers sont dioïques et allogames, et en conséquence des autopollinisations ne peuvent être réalisées.

- Par ailleurs, le peuplier a surtout été choisi pour sa croissance juvénile rapide et la facilité de le manipuler expérimentalement. La capacité de cet arbre à grandir de 3 m par an dans de bonnes conditions permet d'évaluer efficacement les modifications génétiques et il est possible de mesurer à court terme les effets de stress biotiques ou abiotiques appliqués. Par ailleurs, pour certains clones, comme le clone 717-1-B4 (*P. tremula* × *P. alba*), la propagation végétative de manière clonale est très facile en culture *in vitro*. Elle permet de conserver des génotypes intéressants, de s'affranchir de la production de graines, et de réaliser des expériences sur un grand nombre d'échantillons. Enfin, sa culture en serre ou pépinière est également aisée. Par contre, sa grande taille impose des contraintes de place lors des expériences.

- Enfin, sa transformation aisée par *Agrobacterium tumefaciens* a aussi été un argument déterminant pour le choix du peuplier comme arbre modèle pour la biologie intégrative. En effet, une telle facilité de transformation génétique n'est pas le cas pour les autres arbres, notamment l'eucalyptus et les pins. Le clone de peuplier hybride grisard 717-1-B4 (*P. tremula* × *P. alba*) a été l'un des premiers clones utilisé pour la transformation génétique (Leplé *et al.*, 1992) et il est aujourd'hui encore largement utilisé pour l'expression des transgènes. En effet, c'est l'un des clones de peuplier pouvant être facilement transformé par *Agrobacterium* et régénéré efficacement en des arbres transgéniques intacts au bout de 6 à 10 mois (Leplé *et al.*, 1992 ; Jouanin *et al.*, 1993). Malgré cela, tous les génotypes de peuplier ne peuvent être soumis à la transformation génétique par *Agrobacterium* et certaines espèces, comme *P. trichocarpa*, ont un taux de transformation faible (Busov *et al.*, 2005). De plus, le temps de génération est long chez les peupliers. Contrairement à certain eucalyptus et beaucoup de saules pouvant entrer en floraison au bout d'un an ou deux, les peupliers, pour la plupart, ne fleurissent qu'après quatre années de croissance (Bradshaw *et al.*, 2000). C'est un

handicap pour la génétique classique, et la production de mutants transgéniques homozygotes n'est pas envisageable. Toutefois la capacité de produire des peupliers transgéniques hétérozygotes est un atout essentiel dans l'étude fonctionnelle des gènes.

**Tableau 2.** Nombre d'ESTs de peuplier soumises à la base de données Genbank et accessibles sur le site Internet du NCBI au 13/04/07.

<b>Espèce ou hybride</b>	<b>Nombre d'EST</b>
<i>P. trichocarpa</i>	89943
<i>P. tremula</i> × <i>P. tremuloides</i>	76160
<i>P. trichocarpa</i> × <i>P. deltoides</i>	53208
<i>P. tremula</i>	37313
<i>P. trichocarpa</i> × <i>P. nigra</i>	20130
<i>P. nigra</i>	15257
<i>P. deltoides</i>	14661
<i>P. euphratica</i>	13905
<i>P. tremuloides</i>	12813
<i>P. × canescens</i>	10446
<i>P. euramericana</i>	10157
<i>P. alba</i> × <i>P. tremula</i> var. <i>glandulosa</i>	7595
<i>P. fremontii</i> × <i>P. augustifolia</i>	5410
<i>P. tomentiglandulosa</i>	1650
<i>P. tremula</i> × <i>P. alba</i>	585
<i>P. simonii</i>	4
Total	369237

### 1.5. Le génome séquencé de *Populus trichocarpa*

L'utilisation du genre *Populus* comme un système modèle pour la biologie des arbres et des plantes pérennes ligneuses a largement été accompagnée par le développement rapide de ressources en biologie moléculaire et génomique pour le peuplier. En effet, beaucoup d'outils et de ressources génétiques comme des banques BAC, et des marqueurs polymorphes tels que les microsatellites, les RFLPs (restriction fragment length polymorphisms), les AFLPs (amplified fragment length polymorphisms) étaient disponibles avant la publication du génome du peuplier et l'ensemble de ces outils a servi à créer des cartes physiques et génétiques. De plus, un grand nombre d'étiquettes de séquences exprimées (EST), ou d'ADNc pleine longueur ont été séquencés. Le premier jeu d'ESTs publié comptait 5692 séquences issues de banques construites à partir de tissus formant du bois (Sterky *et al.*, 1998). Depuis, de nombreux groupes ont contribué à l'effort de séquençage d'EST en utilisant différentes espèces ou hybrides de peuplier. L'engouement est tel que d'Octobre 2003 jusqu'à

## Introduction Générale

Avril 2007, le nombre de séquences EST a quasiment triplé, pour atteindre un total de 369237 ESTs (Tableau 2). En Avril 2007, il y avait pour le genre *Populus* 465564 séquences nucléotidiques, 2253 séquences de protéines déduites, et 22399 unigènes dans la base de données du « National Center for Biotechnology Information » (NCBI). Les efforts de séquençage d'ESTs n'ont pas été réalisés uniquement à partir du tissu bois, objet d'étude de la plupart des groupes de recherche (Déjardin *et al.*, 2004 ; Sterky *et al.*, 2004), mais aussi à partir de racines (Kohler *et al.*, 2003), de tissus soumis à un stress abiotique comme la sécheresse (Brosche *et al.*, 2005), ou abiotique comme les attaques d'insectes herbivores (Ralph *et al.*, 2006). L'analyse de 102019 ESTs groupées en 24644 unigènes, a montré une grande similarité entre les gènes de peuplier et ceux d'*Arabidopsis* pouvant faciliter les approches de génomique comparée (Sterky *et al.*, 2004). Grâce à ce travail international de séquençage d'ESTs, ainsi que le séquençage d'ADNc pleine longueur (Ralph *et al.*, 2006 ; Tuskan *et al.*, 2006), l'annotation des gènes de peuplier a été grandement facilitée.

Par ailleurs, d'autres outils d'analyses génomiques comme la transcriptomique ont été développés. Après les premières études transcriptomiques réalisées sur des échantillons de bois de peuplier et avec des lames composées de 2995 puis 13526 clones ESTs (Hertzberg *et al.*, 2001) (Schrader *et al.*, 2004b), des lames de microarray comprenant 25000 et 27000 ADNc ont été construites en Suède et en France respectivement (Jansson & Douglas, 2007). De même au Canada, une lame comprenant 15400 ADNc a été utilisée pour l'analyse de la défense du peuplier contre les insectes herbivores (Ralph *et al.*, 2006). Aussi, une puce microarray à oligonucléotide (sondes de 60-mer) représentant l'ensemble des gènes de peuplier a déjà été utilisée pour l'analyse de l'expression des gènes dupliqués (Tuskan *et al.*, 2006). En parallèle, des études protéomiques ont été réalisées pour identifier des protéines présentes dans le bois, dans différents tissus, ou induites par un stress hydrique (Du *et al.*, 2006 ; Plomion *et al.*, 2006), et les analyses métabolomiques commencent à émerger avec la recherche de métabolites dans le bois de tension (Andersson-Gunneras *et al.*, 2006).

Le peuplier est ainsi devenu un arbre modèle pour la génétique moléculaire et la recherche en génomique, et le clone de peuplier « Nisqually 1 » (*Populus trichocarpa*) a été le premier arbre dont le génome est entièrement séquencé (Tuskan *et al.*, 2006). Le séquençage du génome et l'assemblage des séquences ont été terminés au cours de l'année 2006. Il a été choisi pour son génome de taille modeste qui contient  $485 \pm 10$  millions de paires de bases (Mb) réparties sur 38 chromosomes ou 19 groupes de liaisons. Cette taille est comparable à celle du génome de riz (*Oryza sativa*) avec 390 Mb (Sequencing Project International Rice Genome, 2005), correspond à quatre fois celle du génome d'*Arabidopsis* avec 125 Mb (The

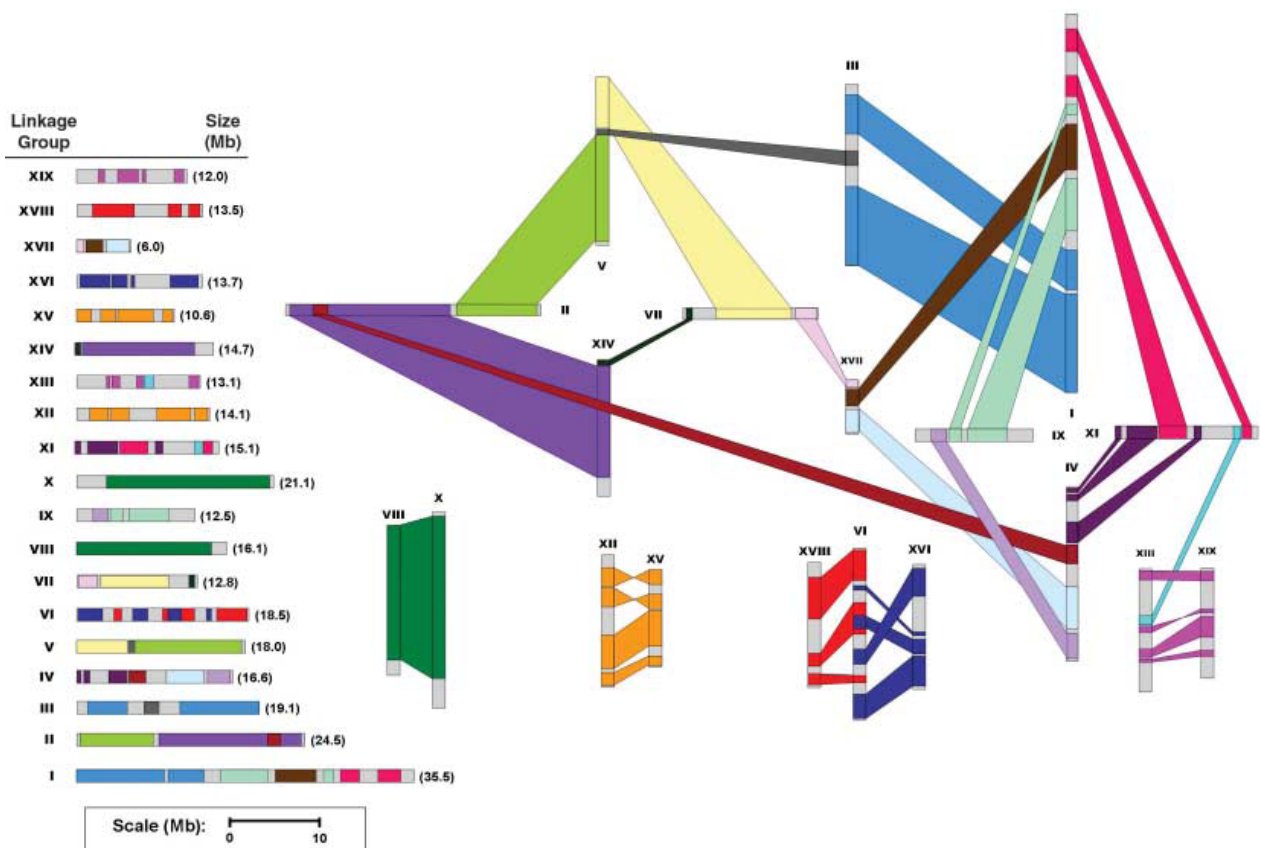
## Introduction Générale

*Arabidopsis* Genome Initiative, 2000), et est 40 à 55 fois inférieure à celle des génomes de pins variant de 20,8 à 26,9 Gb. L'annotation du génome a été réalisée avec une diversité de programme de prédiction de gènes comme « GrailExp6 », ou « Genewise » (Xu & Uberbacher, 1997 ; Birney & Durbin, 2000). De plus, les modèles de gènes des génomes d'*Arabidopsis* et de riz ont facilité cette annotation. Ainsi, 45555 modèles de gène codant pour des protéines putatives ont été identifiés. Le génome du peuplier contient un plus grand nombre de gènes comparé aux deux autres génomes entièrement séquencés de riz et d'*Arabidopsis* contenant approximativement 38000 et 25000 gènes respectivement (The Arabidopsis Genome Initiative, 2000 ; Sequencing Project International Rice Genome, 2005). De plus, 89 % des gènes prédits de peuplier présentent une homologie avec des protéines de la base de donnée non redondante du NCBI, et environ 12 % des modèles de gènes n'ont pas de similarité avec des gènes d'*Arabidopsis* (Tuskan *et al.*, 2006). Par ailleurs, de par les caractères dioïque, anémophile et allogame du genre *Populus*, plus d'un million de polymorphismes simple de nucléotide (SNP) et d'insertion/délétion ont été identifiés au sein du génome hétérozygote de *P. trichocarpa*, avec en moyenne 2,6 sites polymorphes par kilobase. Toutefois, 83 % de ce polymorphisme touche des régions non codantes. Par ailleurs, l'analyse de la séquence génomique du peuplier révèle une duplication récente du génome entier. Elle a affecté approximativement 92 % du génome, et 8000 gènes paralogues fortement similaires entre eux ont été identifiés. Un grand nombre de segments chromosomiques ont un ou plusieurs autres segments paralogues situés ailleurs sur le génome (Fig. 7). La comparaison des gènes orthologues entre les genres *Populus* et *Salix*, ainsi que la colinéarité des cartes génétiques entre ces deux genres suggèrent fortement que cette duplication du génome entier se serait aussi produite chez le genre *Salix* (Tuskan *et al.*, 2006). Cette duplication récente a ainsi été appelée duplication « salicoid ». De plus, la colinéarité des cartes génétiques au sein des espèces du genre *Populus* montrent que ce large réarrangement chromosomique s'est probablement produit avant la divergence des espèces de peuplier, chez un ancêtre commun possédant 10 paires de chromosomes. Un autre événement de duplication plus ancien couvrant environ 59 % du génome de peuplier se serait produit chez l'ancêtre commun du peuplier et d'*Arabidopsis*, avant ou après la spéciation des lignées d'eurosidés I et II (Tuskan *et al.*, 2006). L'expansion de famille de gènes par duplication en tandem semble moins fréquente chez le peuplier que chez *Arabidopsis*. Par ailleurs, la perte de gènes après la duplication « salicoid » a été plus faible que celle s'étant produite après la duplication plus ancienne portée par les deux génomes d'*Arabidopsis* et *Populus* ; si bien qu'un gène simple copie chez *Arabidopsis* se retrouve souvent en double copie chez le peuplier. Enfin, plusieurs



## Introduction Générale

familles de gènes chez le peuplier, telles que celles impliquées dans la biosynthèse des lignines et de la cellulose, ont évolué par un accroissement du nombre de leurs gènes. Beaucoup de gènes dupliqués dans le génome du peuplier ont acquis des expressions spécifiques expliquant leur conservation (Tuskan *et al.*, 2006). L'évolution de la structure des familles de gènes chez les eudicotylédones peut maintenant être examinée avec plus de détails avec la disponibilité des deux génomes d'*Arabidopsis* et de peuplier. Les différences entre ces deux espèces dans la rétention ou la perte de gènes dupliqués au sein des familles de gènes nous aidera à identifier les gènes ayant des fonctions communes aux eudicotylédones et les gènes ayant acquis un rôle spécifique pour l'une ou l'autre lignée.



**Figure 7.** Réorganisation à l'échelle des chromosomes de l'événement de duplication le plus récent affectant le génome entier (Tuskan *et al.*, 2006).

Les couleurs communes réfèrent aux blocs de génomes homologues portés par deux chromosomes et probablement apparus après une duplication de génome chez les *Salicaceae*. Les groupes de liaison sont indiqués par des chiffres romains. Le diagramme de gauche met en évidence la nature chimérique de certains groupes de liaison.

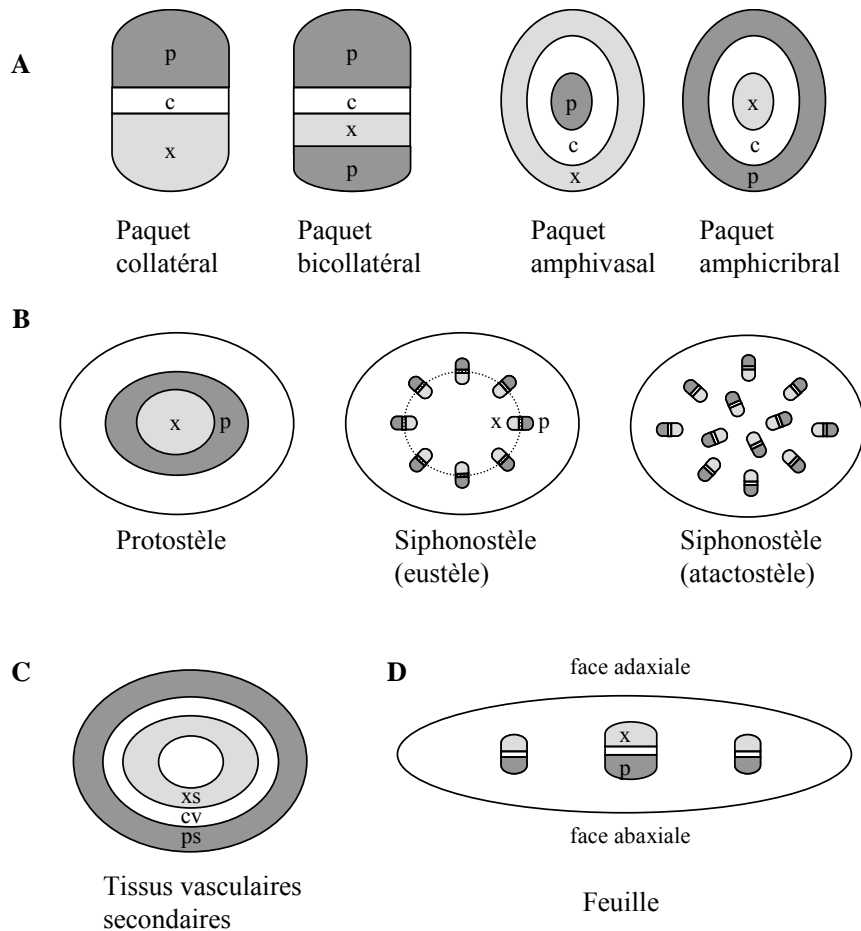
## **2 Aspects physiologique et moléculaire de la formation des tissus vasculaires primaire et secondaire chez les plantes et le peuplier**

### *2.1. La formation du système vasculaire primaire*

Avant de parler du bois, un tissu vasculaire secondaire, il est nécessaire d'évoquer les mécanismes de formation des tissus vasculaires primaires intervenant dans toutes les plantes supérieures, y compris les arbres. Les tissus vasculaires sont composés de deux unités basiques : le xylème et le phloème. Le xylème transporte, à travers la sève brute, l'eau, les minéraux et des hormones tels que les cytokinines et l'acide abscissique. Il stocke l'eau et les nutriments et permet le soutien mécanique du poids de la plante. Le phloème conduit la sève élaborée pour la distribution des produits de la photosynthèse tel que le saccharose. Il permet la translocation des protéines et ARNm impliqués dans la croissance et le développement de la plante (Ye, 2002). Chez *Arabidopsis*, la formation des tissus vasculaires primaires commence durant l'embryogenèse le long de l'axe hypocotyle/racine, avec la différenciation en fibres des cellules méristématiques primaires du procambium vasculaire. Après la germination lors de la croissance primaire des tiges et des racines, les cellules du procambium se différencient pour former à la fois le phloème primaire et le xylème primaire, le tout confiné dans des paquets ou faisceaux vasculaires. Le xylème primaire est constitué des éléments de trachéide conduisant la sève brute, de fibres du xylème, et de cellules du parenchyme du xylème. Chez les angiospermes, les trachéides sont appelés éléments de vaisseaux et sont perforés à chaque extrémité pour former une colonne creuse continue appelée vaisseau. Le phloème est composé d'éléments criblés conduisant la sève élaborée, de cellules compagnes du parenchyme, et de fibres. Les éléments criblés chez les plantes à fleurs sont des tubes criblés qui sont connectés à travers des plaques criblées pour former une colonne continue (Ye, 2002).

Il y a une grande plasticité d'organisation des tissus vasculaires au sein d'un paquet vasculaire, et l'organisation vasculaire la plus commune est un placement parallèle du xylème vers l'intérieur et du phloème vers l'extérieur (Ye, 2002 ; Nieminen *et al.*, 2004). On parle alors de paquet vasculaire collatéral (Fig. 8A). Mais nous pouvons rencontrer d'autres organisations, comme par exemple chez les *Solanaceae* où le xylème est placé en parallèle avec un phloème externe et un phloème interne formant un paquet vasculaire bicollatéral (Ye, 2002).

## Introduction Générale



**Figure 8.** Représentation schématique des différentes organisations de tissus vasculaires en coupe transversale (Ye, 2002)

(A) Organisation des tissus vasculaires dans les paquets vasculaires primaires. Le procambium (c) est localisé entre le xylème et le phloème. Le paquet vasculaire collatéral a un xylème (x) et un phloème (p) placés en parallèle. Le paquet vasculaire bicollatéral a un xylème placé entre du phloème externe et interne. Dans le paquet vasculaire amphivasal le xylème entoure le phloème, alors que dans le paquet vasculaire amphicribral le phloème entoure le xylème. (B) Organisation des paquets vasculaires dans la tige. La protostèle a un xylème localisé au centre et un phloème entourant le xylème. Dans la siphonostèle, les paquets vasculaires sont organisés en cercle (eustèle) ou dispersés partout dans la masse tissulaire (atactostèle). Le xylème est toujours placé vers le centre et le phloème vers la périphérie. (C) Organisation des organes vasculaires en croissance secondaire. Le cambium vasculaire (cv) forme un cercle, avec de part et d'autre du xylème secondaire (xs) vers l'intérieur et du phloème secondaire (ps) vers l'extérieur. (D) Paquets vasculaires collatéraux dans la feuille. Le xylème est localisé vers la face adaxiale de la feuille et le phloème est localisé vers la face abaxiale.

De la même façon, les plantes ont élaboré diverses organisations pour le placement des paquets vasculaires au sein de la tige ou de la racine (Fig. 8B). Chez beaucoup de dicotylédones, le xylème forme une masse solide entourée de phloème au centre de la racine primaire ou protostèle. Les paquets vasculaires peuvent à l'inverse être arrangés individuellement avec deux modèles majeurs dans la tige alors appelée siphonostèle. Dans le premier modèle (eustèle), les paquets vasculaires individuels sont organisés en cercle dans la tige ou la racine. C'est le cas de beaucoup de dicotylédones notamment pour les tissus primaires de peuplier, et les racines de monocotylédones. Dans l'autre modèle (atactostèle),

les paquets vasculaires sont éparpillés d'un bout à l'autre des tissus, comme cela est observé dans les tiges de monocotylédones (Ye, 2002). Dans les feuilles, les paquets vasculaires sont appelé veines, avec généralement chez les dicotylédones une veine principale au milieu et un réseau de veines mineures appelées nervures. Les paquets vasculaires des feuilles sont collatéraux avec un phloème à la face abaxiale et un xylème à la face adaxiale de la feuille (Fig. 8D). Dans les jeunes tiges de peupliers (*Populus tremula* × *P. alba*) âgés d'un mois, la zone de croissance primaire se situe directement sous l'apex où la tige s'allonge et est encore verte (Prassinos *et al.*, 2005). De plus, la zone de transition entre les croissances primaire et secondaire se situe entre les quatrième et sixième nœuds, et la zone de croissance secondaire est localisée vers la base de la tige où la formation de bois est bien visible. Au niveau de la zone de transition quand la tige devient mature, le procambium donne le cambium fasciculaire et du cambium interfasciculaire commence à se développer entre les paquets vasculaires pour finalement former un cylindre continu appelé cambium vasculaire (Fig. 8C) (Ye, 2002 ; Nieminen *et al.*, 2004). La plante annuelle *Arabidopsis* est capable de produire dans certaines conditions du xylème secondaire à la base des inflorescences, et est aussi utilisée comme modèle dans l'étude de la formation du bois.

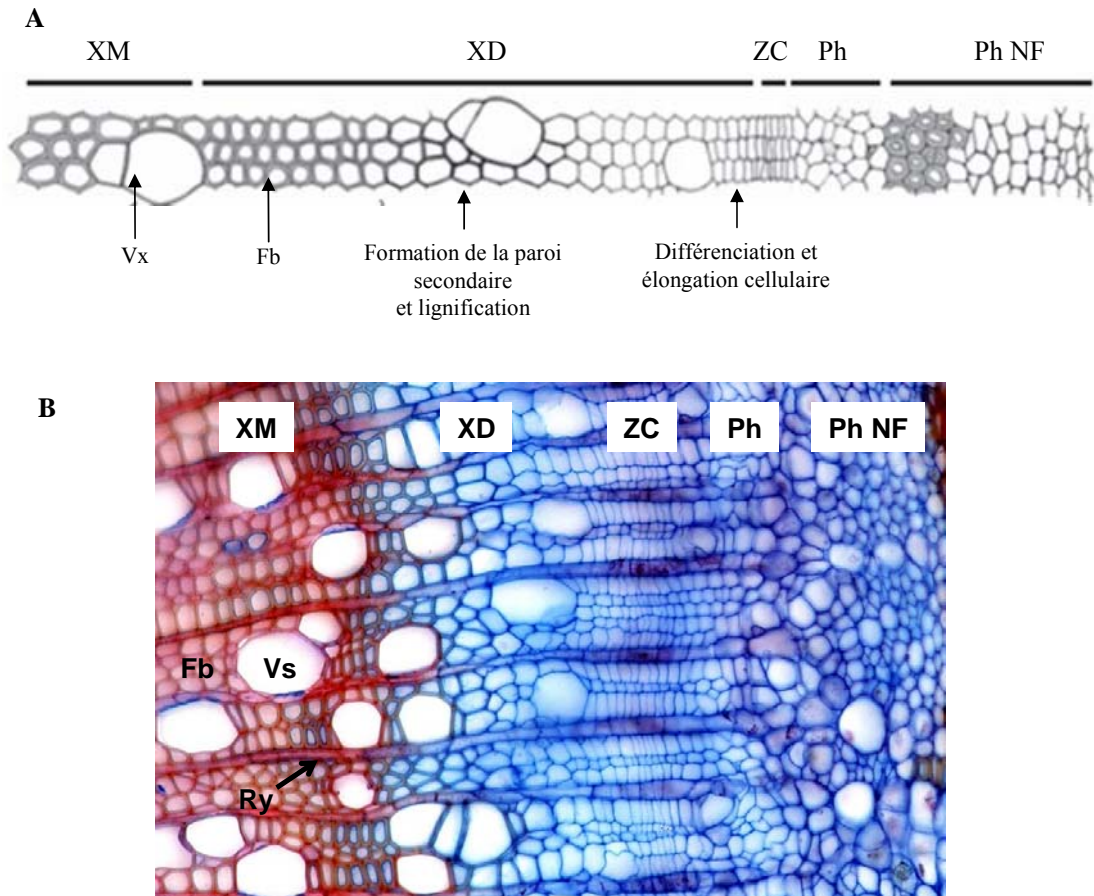
### 2.2. La formation du bois ou xylogenèse chez le peuplier

Le bois ou xylème secondaire est un tissu hétérogène et complexe, si bien qu'au sein d'une même plante, il existe différentes sortes de bois : le bois juvénile, le bois mature, le bois de printemps, le bois d'été, le bois de cœur, et le bois de réaction appelé bois de tension chez les angiospermes (Mellerowicz *et al.*, 2001 ; Plomion *et al.*, 2001). Chez les arbres, le bois a différentes fonctions : le soutien mécanique permettant aux arbres de se développer durant plusieurs années, le transport de la sève brute des racines aux feuilles sur de longues distances, et enfin le stockage d'éléments nutritionnels notamment pendant l'hiver. Les conifères possèdent deux types de cellules : les cellules radiales assurant le transport de sève entre le phloème et le xylème, et les trachéides connectées entre elles par des ponctuations aréolées et assurant à la fois le rôle de transport de la sève brute et le rôle de soutien mécanique (Plomion *et al.*, 2001). Contrairement aux conifères, le bois des feuillus est hétérogène avec quatre types cellulaires principaux : les fibres pour le soutien mécanique, les vaisseaux pour la conduction de la sève, les cellules axiales parenchymateuses et les cellules de parenchyme des rayons pour l'accumulation des réserves et le transport radial de l'eau et

des nutriments vers le centre de la tige ou du tronc. Au niveau des racines l'anatomie du bois ou xylème secondaire est similaire à celle des tiges ou du tronc (Chaffey *et al.*, 2002). Le bois est un tissu secondaire issu de l'activité d'un méristème secondaire appelé cambium vasculaire permettant la croissance en largeur du tronc ou des tiges. Ce cambium secondaire situé entre le phloème et le xylème forme un cercle continu au sein du tronc ou des tiges (Fig. 8C). La formation du bois ou xylogenèse est un processus complexe faisant l'objet de nombreuses études et qui peut être résumé très succinctement par les quatre processus physiologiques suivants : division cellulaire à partir de cellules méristématiques du cambium, expansion cellulaire, dépôt et lignification de la paroi secondaire, et mort cellulaire programmée (Fig. 9). Au cours de leur différenciation, les cellules du bois se renforcent par l'épaississement et l'imperméabilisation de leurs parois cellulaires.

### 2.2.1. La zone cambiale, une région méristématique de division et d'expansion cellulaire

Le cambium, ou assise génératrice libéro-ligneuse, est une zone méristématique secondaire initiant la xylogenèse et qui assure la pérennité des arbres (Lachaud *et al.*, 1999). Cette assise cylindrique constituée de cellules méristématiques organisées en files radiales génère en effet, par divisions périclines le phloème secondaire ou liber côté extérieur et le xylème secondaire ou bois côté intérieur de la tige (Fig. 9 et 10). Contrairement aux autres méristèmes, le cambium est un tissu complexe constitué de deux types cellulaires distincts morphologiquement, à parois fines et très vacuolisées, appelés initiales fusiformes et initiales radiales (Mellerowicz *et al.*, 2001 ; Chaffey *et al.*, 2002). L'identité de ces cellules cambiales est déterminée par des signaux positionnels plutôt que par la lignée cellulaire, puisque l'interconversion entre les initiales fusiformes et les initiales radiales est un phénomène commun (Larson, 1994). La zone cambiale chez *P. tremula* × *P. tremuloides* est constitué d'environ 7-10 cellules fusiformes durant la phase de croissance active (Chaffey *et al.*, 2002). Les initiales fusiformes sont allongées de façon axiale, et les initiales radiales isodiamétriques sont plus rares et regroupées en îlots lenticulaires. Les cellules de la zone cambiale ont une paroi primaire fine appelée lamelle moyenne, composée de microfibrilles de cellulose arrangées au hasard durant la période active.



**Figure 9.** Les différentes étapes de la formation du bois chez le peuplier

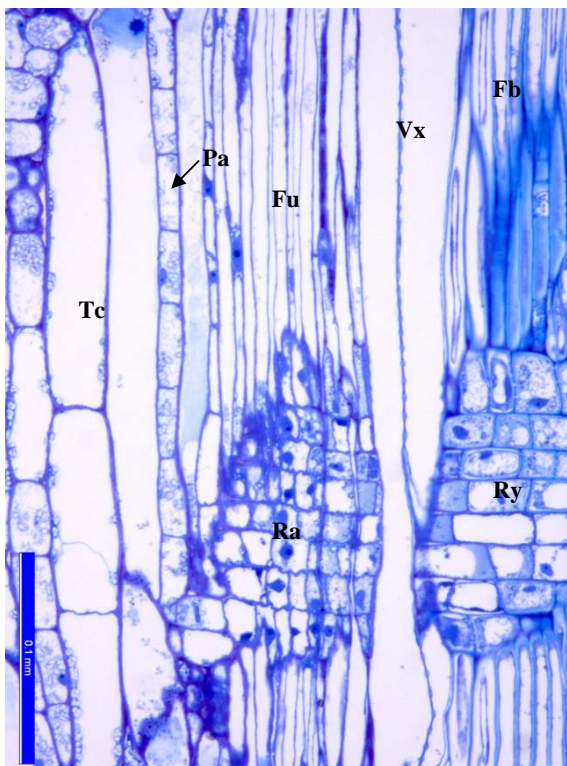
(A) Représentation schématique de la différenciation du xylème chez le peuplier d'après (Israelsson *et al.*, 2005).

(B) Coupe transversale de tige de peuplier colorée à la safranine bleu-astra et observée en microscopie optique (Photo réalisée par Laurans F., INRA Orléans). La cellulose est colorée en bleu, tandis que la lignine est colorée en rouge.

Les cellules de la zone cambiale (ZC) donnent naissance aux cellules du phloème (Ph) coté externe de la tige. Le phloème produit par la suite du phloème non fonctionnel ou cortex (Ph NF). Coté interne, elles donnent naissance aux cellules du xylème qui se différencient (XD) pour donner au final le xylème mature (XM) lignifié. Le xylème est composé de trois types cellulaires principaux : les fibres (Fb), les vaisseaux (Vs), et les rayons (Ry).

Les divisions périclines des initiales fusiformes cambiales sont très particulières puisque ces cellules se divisent par leur bord longitudinal. La cloison de paroi primaire devant être formée pour chaque division péricline d'une initiale fusiforme est la plus grande possible au sein de la cellule, alors que pour d'autres types cellulaires cette cloison de paroi primaire générant deux cellules filles est habituellement la plus petite possible (Mellerowicz *et al.*, 2001). La biosynthèse de la plaque cellulaire, de la lamelle moyenne, de la membrane plasmique et de la paroi primaire est donc quantitativement plus importante que pour d'autres cellules. De plus le phragmoplaste, une matrice riche en polysaccharides déposée entre les cellules filles durant la division cellulaire et générant ensuite la plaque cellulaire, doit traverser la grande vacuole centrale pour atteindre les deux extrémités de la cellule. Par

ailleurs, des divisions anticlinales, où la nouvelle paroi cellulaire formée est placée radialement, interviennent aussi au niveau de ces cellules cambiales pour maintenir ou accroître la circonférence du cambium au cours de la croissance (Mellerowicz *et al.*, 2001). Les divisions anticlinales jouent un rôle important dans la régulation de la longueur des initiales fusiformes, et indirectement dans la longueur des fibres de xylème (Larson, 1994). Lors de la division anticlinale, les cellules filles sont plus courtes que leur cellule mère, de par l'orientation oblique de la nouvelle paroi radiale formée. La longueur de ces initiales fusiformes augmente ensuite graduellement par de la croissance intrusive à leurs extrémités, jusqu'à la prochaine division anticlinale. Cette activité cambiale varie au cours des saisons et notamment connaît une période de dormance pendant l'hiver. Le cambium reste actif pendant des dizaines d'années, et il est admis qu'il est contrôlé par des circuits internes pouvant être analogues à ceux des méristèmes apicaux (Groover, 2005 ; Groover *et al.*, 2006).



**Figure 10.** La différenciation des initiales cambiales en cellules du phloème et du xylème dans du bois « normal » de peuplier (*Populus tremula* × *P. alba*)

Coupe longitudinale radiale de la zone cambiale colorée au bleu de méthylène azur2 (Photos Laurans F.). Les initiales fusiformes (**Fu**) se différencient pour produire vers le xylème les éléments de vaisseaux (**Vx**) et les fibres (**Fb**), et vers le phloème les tubes criblés (**Tc**) et les cellules parenchymateuses du phloème (**Pa**). Les initiales radiales (**Ra**) d'aspect cubique donnent naissance aux cellules des rayons (**Ry**) en forme de brique. Nous pouvons remarquer que la croissance cellulaire est très précoce pendant la différenciation du xylème et du phloème, et les éléments de vaisseaux ont une croissance radiale plus marquée que les fibres du xylème.

Chez le peuplier, quatre principales formes cellulaires sont générées coté xylème à partir de ces deux types de cellules cambiales (Mellerowicz *et al.*, 2001 ; Chaffey *et al.*, 2002). Coté xylème, les initiales fusiformes donnent naissance aux éléments de vaisseaux conduisant la sève brute, aux fibres libriformes assurant le soutien mécanique de l'arbre (Fig. 9 et 10) et enfin aux cellules axiales parenchymateuses. Les éléments de vaisseaux sont soit individuels, soit regroupés avec au moins deux cellules adjacentes en file radiale, ou en

paquet de plusieurs éléments de vaisseaux. Coté phloème, les initiales fusiformes engendrent les tubes criblés ou fibres de phloème. Les initiales radiales d'aspect cubique produisent les cellules de rayon du xylème et du phloème qui prennent une forme de brique (Fig. 10). Ces cellules sont plus tard distinguées en cellules de contacts et cellules parenchymateuses des rayons. Cette distinction est basée respectivement sur leurs associations ou non avec les éléments des vaisseaux. Ainsi, les initiales radiales sont à l'origine des rayons qui permettent la communication entre le xylème et le phloème. La production de cellules du xylème déplace le cambium de manière centrifuge vers l'extérieur. L'établissement de la destinée cellulaire et la manière dont ces cellules se différencient restent encore un mystère du processus de xylogenèse (Mellerowicz *et al.*, 2001).

### 2.2.2. L'expansion cellulaire et l'élaboration de la paroi cellulaire primaire

Après avoir perdu leur capacité de division, les cellules mères du xylème vont connaître une phase d'expansion cellulaire (Fig. 10). Toutes les cellules dans la zone d'élongation jouxtant le cambium subissent un élargissement. Chaque type cellulaire subit des modifications qui lui sont propres, et différentes selon l'étendue, la polarité et le type d'élargissement (Mellerowicz *et al.*, 2001). Le diamètre des vaisseaux augmente considérablement et leur longueur axiale/longitudinale varie très peu, alors que les fibres ont une expansion longitudinale très marquée et sont légèrement plus larges que les initiales fusiformes. Les fibres atteignent une longueur d'environ 0,4 à 0,6 mm chez le peuplier et leur longueur augmente avec l'âge de l'arbre (Larson, 1994 ; Jourez *et al.*, 2001). Les vaisseaux ont une longueur finale moyenne de 0,38 mm. Les cellules axiales parenchymateuses formées tardivement dans la saison sont plus larges mais plus courtes que les initiales fusiformes (Chaffey *et al.*, 2002). Concernant les cellules des rayons, elles s'allongent selon l'axe radial du xylème et prennent une forme de brique. Bien que les fibres présentent une croissance cellulaire anisotrope, la nature exacte du type d'expansion cellulaire n'est pas tranchée. Beaucoup de chercheurs pensent que la croissance des fibres est intrusive aux extrémités de la cellule et nécessite la biosynthèse locale de paroi cellulaire (Wenham & Cusick, 1975 ; Mellerowicz *et al.*, 2001). Cette croissance serait alors similaire à celle localisée aux extrémités du tube pollinique ou des poils racinaires. D'autres, parce que les cellules en développement cessent de s'allonger quand la paroi secondaire commence à être déposée, penchent plutôt pour une croissance diffuse similaire aux cellules non vasculaires (Ye, 2002).



## Introduction Générale

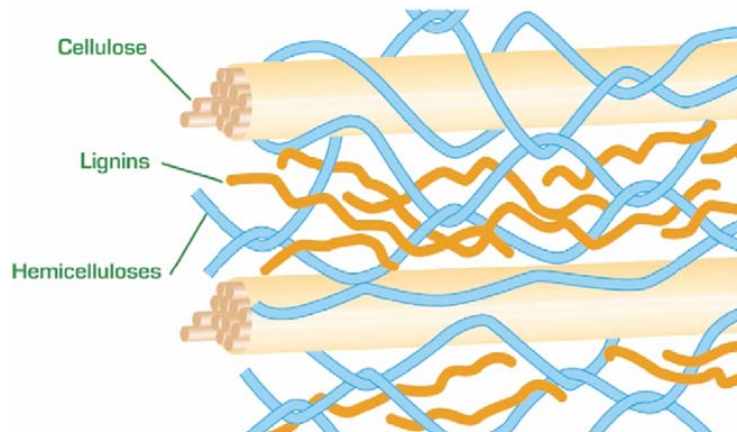
Toutefois, sans exclure l'existence d'une croissance diffuse au niveau des parois latérales, il y a un large consensus au sein de la communauté scientifique supportant l'existence d'une croissance intrusive aux deux extrémités des fibres et des éléments de vaisseaux. Au cours de la croissance secondaire du cambium et de la formation des éléments de vaisseaux et des fibres du xylème secondaire, ces deux types de croissance diffuse et intrusive semblent donc se produire simultanément. Les modifications de taille des cellules de xylème (fibres, vaisseaux et rayons) sont concomitantes et liées à un dépôt de paroi primaire. La croissance des cellules se termine dès que la paroi secondaire commence à se former.

La paroi cellulaire primaire est un constituant essentiel aux plantes, mais aussi à d'autres organismes procaryotes et eucaryotes tels que les bactéries, levures, et champignons. La paroi cellulaire détermine notamment la morphologie des cellules de plantes, peut être considérée comme un exosquelette et permet le maintien de la pression de turgescence des cellules. Son rôle principal est mécanique, mais elle peut avoir d'autres fonctions comme par exemple la signalisation cellulaire, ou la protection des cellules contre les attaques de pathogènes (Cosgrove, 1997). La paroi primaire est mise en place dans les tissus vasculaires secondaires avant le dépôt de la paroi secondaire (Mellerowicz *et al.*, 2001). Sa formation débute au niveau des initiales fusiformes et radiales de la zone cambiale et se poursuit lors de l'expansion cellulaire. Tout d'abord, la lamelle moyenne (0,5 à 1,5  $\mu\text{m}$ ) se forme à partir du phragmoplaste. Elle est en majorité constituée de pectines et assure à la fois l'adhésion et la séparation entre les cellules. Après mise en place de la lamelle moyenne tout de suite après la division cellulaire, la paroi primaire est déposée. D'épaisseur faible (0,1  $\mu\text{m}$ ) et très hydratée (composée aux deux tiers d'eau), la paroi primaire a des propriétés élastiques qui permettent l'expansion cellulaire. La paroi primaire des cellules végétales est constituée en moyenne de 30 % de celluloses, 30 % d'hémicelluloses, 35 % de pectines et de 1 à 5 % de protéines pariétales (Cosgrove, 1997). La teneur et la nature des constituants de la paroi primaire varie en fonction du type cellulaire et des espèces. Pour information, la composition de la paroi primaire dans le xylème de peuplier est indiquée dans le tableau 3 (Mellerowicz *et al.*, 2001). La cellulose est déposée de manière aléatoire et les différentes couches de microfibrilles de celluloses sont orientées au hasard dans la paroi primaire. Les microfibrilles de celluloses enchevêtrées dans la matrice d'hémicelluloses et de pectines forme un maillage lâche (Kohorn, 2000). Alors que la cellulose est synthétisée au niveau de la membrane plasmique au niveau des complexes de cellulose synthases, les hémicelluloses et pectines sont sécrétées dans l'espace extracellulaire par un système de vésicules de sécrétion. Les pectines forment

une matrice gélifiante qui s'intercale entre les microfibrilles de cellulose et les hémicelluloses (Kohorn, 2000). Plus tard, les polymères de lignines sont incorporés dans la paroi primaire.

**Tableau 3.** Composition biochimique des parois cellulaires primaire et secondaire de xylème de peuplier (Mellerowicz *et al.*, 2001)

Stades de développement du xylème	Composants (% de poids sec)
Formation de la paroi primaire	Pectines (47 %)
	Cellulose (22 %)
	Xylanes (11 %)
	Xyloglucanes (6 %)
	Glucomannanes (1 %)
	Protéines (10 %)
Bois mature : (paroi secondaire, primaire et lamelle moyenne)	Cellulose (43-48 %)
	Xylanes (18-28 %)
	Glucomannanes (5 %)
	Pectines et xyloglucanes (3 %)
	Lignines (19-21 %)

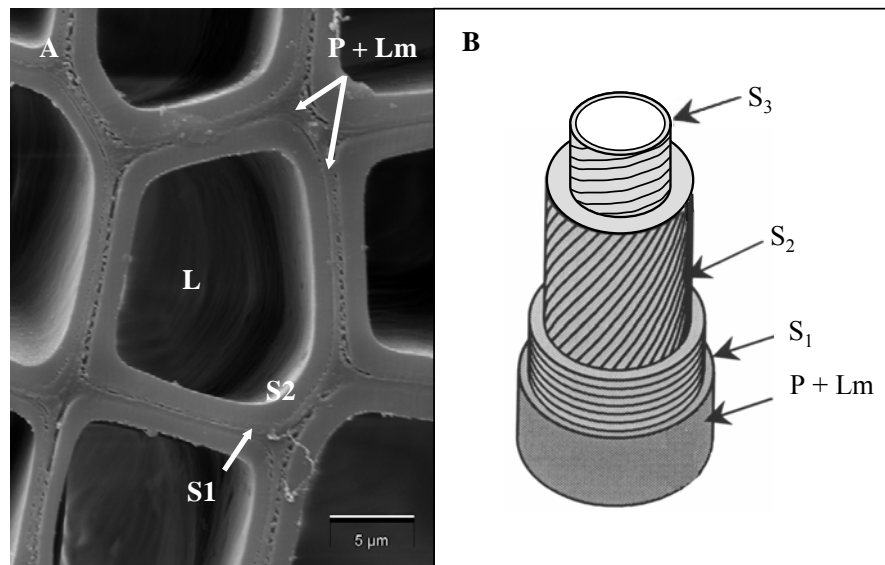


**Figure 11.** Représentation schématique de la paroi secondaire lignifiée. (Boudet *et al.*, 2003)  
Les polymères de lignine sont déposés dans la paroi secondaire entre la matrice d'hémicelluloses et de microfibrilles de cellulose. La lignine forme ainsi un « mastic » colmatant les polysaccharides.

### 2.2.3. La formation de la paroi secondaire

Ayant atteint leur taille définitive, chaque type cellulaire va former une paroi secondaire composée majoritairement de celluloses, d'hémicelluloses, de pectines, de protéines et d'une quantité importante de lignines. La présence de lignines est facilement détectable dans les cellules de bois après coloration des coupes histologiques à la safranine bleu-astra colorant la lignine en rouge et la cellulose en bleu (Fig. 9). La composition approximative de ces polymères dans la paroi secondaire est de 45 % de cellulose, 25 % d'hémicellulose et 20 % de lignine (Timell, 1969). La paroi secondaire est une structure

formée d'une matrice de lignines et d'hémicellulose renforcée par des câbles de microfibrilles de cellulose (Fig. 11) (Boudet *et al.*, 2003). Elle est déposée à l'intérieur de la paroi primaire, et contrairement à cette dernière la paroi secondaire est plus résistante mécaniquement. La composition biochimique de la paroi secondaire est nettement différente de celle de la paroi primaire chez le peuplier (Tableau 3). Comparée à la paroi primaire, la paroi secondaire contient deux fois plus de cellulose, et la teneur en xylanes augmente également. Une caractéristique importante de la paroi secondaire est la diminution drastique de pectine et la présence de 20 % de lignine (Mellerowicz *et al.*, 2001). De plus, de par la densification du réseau de microfibrilles de cellulose et de l'incorporation de polymères de lignine à caractère hydrophobe, le contenu en eau diminue fortement dans la paroi secondaire. Alors que la paroi primaire est composée de microfibrilles de cellulose ayant une disposition totalement désordonnée, les microfibrilles dans les différentes couches de la paroi secondaire prennent une disposition très ordonnée et sont parallèles les unes aux autres. La paroi, apposée sur la face interne de la cellule est formée de trois couches S1, S2 et S3 (Fig. 12). Ces dernières diffèrent principalement par leur épaisseur et l'angle des microfibrilles de cellulose.



**Figure 12.** Les fibres de bois normal chez le peuplier.

(A) Les fibres de xylème de bois normal de peuplier en microscopie électronique à balayage (Photo Laurans F., INRA Orléans). Ces fibres présentent une couche **S2** épaisse, tandis que la couche **S1**, plus fine, est difficilement visible. **L** représente la lumière de la fibre, dépourvue de contenu cellulaire. Entre les cellules et dans les coins cellulaires, on trouve la paroi primaire (**P**) et la lamelle moyenne (**Lm**) difficilement distinguables.

(B) Représentation schématique d'une fibre de xylème avec les différentes couches de la paroi et l'orientation des fibres de cellulose. De l'extérieur vers l'intérieur, on a la lamelle moyenne (**Lm**), la paroi primaire (**P**), et la paroi secondaire composée de la succession des couches **S1**, **S2** et **S3**.

## Introduction Générale

La couche S1 la plus externe, est la première synthétisée et correspond à une couche de transition entre la paroi primaire et les couches S2 et S3 plus profondes. C'est la couche la plus mince avec une épaisseur de 0,1 à 0,35  $\mu\text{m}$  et elle ne représente que 10 à 15 % de l'épaisseur totale de la paroi secondaire (Plomion *et al.*, 2001). Cette couche S1 possède des microfibrilles avec un angle de 60 à 80° par rapport à l'axe longitudinal de la cellule. La couche S2 représente 75 à 85 % de la paroi secondaire avec une épaisseur de 1 à 10  $\mu\text{m}$ . Les microfibrilles de cellulose sont disposées presque parallèlement à l'axe longitudinal de la cellule et forment un angle de 10 à 30° avec l'axe de la cellule. Cette strate S2 joue un rôle prépondérant dans le soutien mécanique de la paroi des cellules lignifiées. Les éléments de vaisseaux du xylème possèdent une couche S2 plus fine que celle des fibres. La couche S3 est la couche la plus interne. Avec une épaisseur de 0,5 à 1,1  $\mu\text{m}$ , elle est relativement fine. L'angle des microfibrilles de cellulose est de 60 à 90° par rapport à l'axe de la cellule et les microfibrilles se trouvent disposées radialement.

Les cellules des vaisseaux à maturité ont une épaisse paroi cellulaire lignifiée et possèdent des perforations simples et larges à chaque extrémité de la cellule. Les perforations simples sont dépourvues de cellulose, ne développent pas de paroi secondaire lignifiée et seront éliminées plus tard pour permettre le transport de sève brute sur de longues distances (Chaffey *et al.*, 2002). De plus, les parois longitudinales des éléments de vaisseaux comportent à la fois des ponctuations aréolées entre les vaisseaux et des ponctuations de contact (non aréolées) de large diamètre entre les vaisseaux et les rayons. Les cellules des rayons ont également une paroi épaisse et lignifiée. Au contact des éléments de vaisseaux, les cellules de contact des rayons portent dans leurs parois radiales les ponctuations de contact. Le diamètre des ponctuations aréolées diminue avec la différenciation des vaisseaux laissant une petite ouverture elliptique, alors que le diamètre des ponctuations de contact ne varie pas au cours de la formation du bois. Les contacts avec les autres cellules des rayons par les parois tangentielle et transverses se font par des ponctuations simples. Les fibres et les cellules axiales parenchymateuses présentent des ponctuations simples dans leurs parois. Les ponctuations des fibres ont été appelées ponctuations simples car leur diamètre est beaucoup plus petit que celui des ponctuations aréolées entre les éléments de vaisseaux et des ponctuations de contact entre les vaisseaux et les rayons. Les plasmodesmes sont persistants au sein des ponctuations simples des fibres et des cellules axiales parenchymateuse, mais sont absents des ponctuations aréolées (Chaffey *et al.*, 2002). Ces ponctuations ne contenant pas de paroi secondaire, permettent la communication entre les cellules et le transport d'eau, d'éléments minéraux et de nutriments par la voie symplasmique.

### 2.2.4. L'étape finale : la mort cellulaire programmée

Le dernier événement est la mort cellulaire programmée génétiquement, consécutif à une autolyse des cellules. Cette dernière serait unique aux cellules vasculaires de plantes supérieures et ne ressemblerait pas à l'apoptose, un système de mort cellulaire programmé fréquemment observé chez les animaux où se produit un rétrécissement des cellules et des noyaux, une fragmentation de l'ADN et la formation de corps apoptotiques (Fukuda, 1996 ; Turner *et al.*, 2007). La mort cellulaire programmée apparaît chez les fibres du xylème et les éléments de vaisseaux lorsque la lignification est complète. La membrane vacuolaire se collapse et se dissocie, et les enzymes hydrolytiques vacuolaires tels que les protéases, DNAses, et RNAses sont ainsi libérées, hydrolysant le contenu cellulaire (Fukuda, 1996). La dissociation de la membrane vacuolaire provoque une acidification du cytoplasme qui pourrait aussi induire l'activation d'enzymes hydrolytiques cytoplasmiques. Le noyau se condense pour disparaître lui aussi et les cellules se vident de leur contenu par une autolyse des organites cellulaires tels que les plastes, mitochondries et réticulum endoplasmique. Cette accumulation de cellules mortes, principalement les fibres et les vaisseaux constitue le bois. Les cellules de rayons possédant une paroi secondaire lignifiée, restent vivantes longtemps, n'engagent pas de mort cellulaire programmée et assurent une fonction de stockage. Cette mort cellulaire programmée semble donc être régulée indépendamment de la formation de la paroi secondaire et de sa lignification (Mellerowicz *et al.*, 2001).

### 2.3. Rôle du cytosquelette dans la formation du bois

Les microtubules corticaux et les filaments d'actines ont un rôle fondamental dans la croissance cellulaire et l'élaboration des parois cellulaires primaires et secondaires chez les plantes. Les microfilaments d'actines sont importants dans le transport des vésicules de golgi contenant les éléments servant à la synthèse de la paroi (Hepler *et al.*, 2001). Ils pourraient aussi participer à la réception et la transduction de stimuli mécaniques ou chimiques. Les microtubules ont eux une fonction dans la biosynthèse de cellulose et guideraient l'orientation des microfibrilles de cellulose (Paredez *et al.*, 2006). Ils sont également essentiels lors de la cytokinèse pendant la division cellulaire et participent aussi au transport de protéines vers la membrane plasmique. De nombreuses études sur le cytosquelette ont permis de montrer son rôle prépondérant dans les premières phases de différenciation du xylème ainsi que dans la

formation de la paroi secondaire chez les arbres (Chaffey *et al.*, 1999 ; Chaffey *et al.*, 2000a ; Chaffey *et al.*, 2002). L'orientation des microtubules et des microfilaments d'actine est particulière dans les cellules du bois et ces deux éléments du cytosquelette ne sont pas organisés de la même manière. De plus, l'organisation des réseaux de microtubules et de microfilaments d'actine varie dans le xylème secondaire en fonction du type cellulaire, notamment entre les fibres et les cellules de rayons.

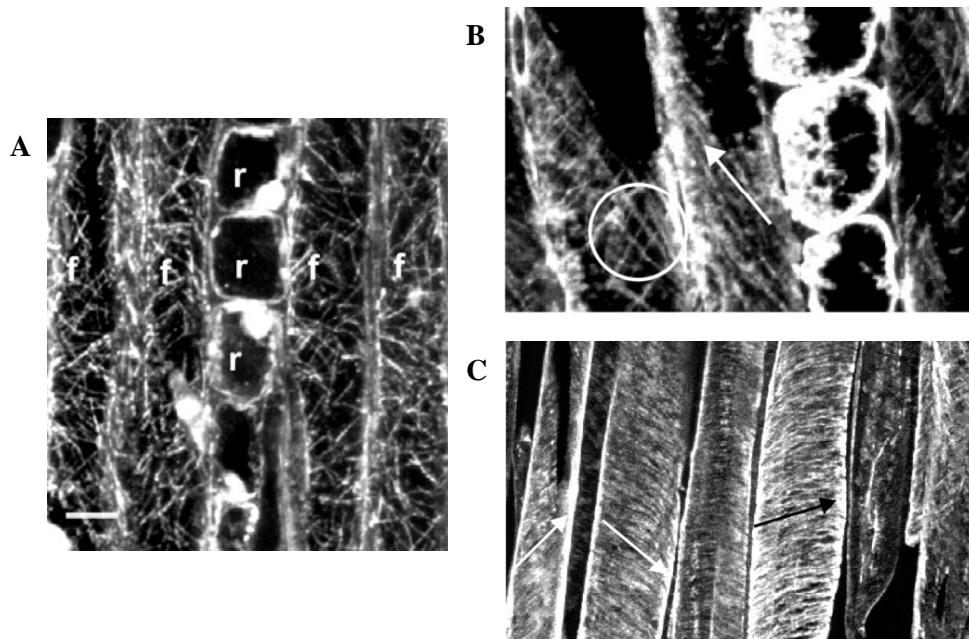
### 2.3.1. Les microtubules

Durant le développement du système vasculaire secondaire chez les arbres, les microtubules passent d'un état totalement désordonné dans les initiales fusiformes à un arrangement plutôt longitudinal et hélicoïdal dans les fibres formant la paroi secondaire. Chez le marronnier (*Aesculus hippocastanum*), et le peuplier (*Populus tremula* × *P. tremuloides*) les microtubules simples et souvent chevauchant sont arrangés au hasard au sein des initiales fusiformes et radiales le long du cytoplasme pariétal (Fig. 13A) (Chaffey *et al.*, 1999 ; Chaffey *et al.*, 2002). Cette orientation des microtubules est consistante avec l'orientation désordonnée des microfibrilles de cellulose de la paroi primaire des cellules cambiales et suggère une influence des microtubules sur l'organisation des microfibrilles de cellulose.

Lors du développement des fibres et des cellules axiales parenchymateuses, dès le début de la phase d'élongation cellulaire et de synthèse de la paroi primaire, les microtubules prennent une organisation plus ordonnée. Les microtubules corticaux relativement clairsemés sont arrangés hélicoïdalement et parallèlement sur toute la longueur des cellules et sont orienté de façon oblique par rapport à l'axe longitudinal des cellules (Fig. 13B) (Chaffey *et al.*, 1999 ; Chaffey *et al.*, 2002). Lors de cette phase d'expansion cellulaire, les microtubules ne sont pas liés au plasmalemme et peuvent s'entrecroiser occasionnellement (Chaffey *et al.*, 2002).

Au cours de l'épaississement de la paroi cellulaire des fibres lors du dépôt de paroi secondaire, le réseau cortical de microtubule orienté de façon hélicoïdale et parallèle est plus dense (Fig. 13C). Les microtubules sont parfois en contact avec le plasmalemme, ne s'entrecroisent pas, et de nombreuses vésicules quelquefois associées aux microtubules sont observées dans l'espace périplasmique (Chaffey *et al.*, 1997a ; Chaffey *et al.*, 1999). De plus, dans les fibres normales en différenciation, l'angle entre les hélices de microtubules et l'axe longitudinal de la cellule varie entre les différentes fibres et l'orientation des microtubules

peut éventuellement devenir transverse. Ces changements d'orientation des microtubules corticaux reflètent les changements d'orientation des microfibrilles de cellulose au cours de l'expansion cellulaire et de la formation des différentes couches S1, S2 et S3 de la paroi secondaire des fibres. En effet, un parallélisme est observé entre les microtubules et les striations des microfibrilles de cellulose (Chaffey *et al.*, 2002). Il en est de même chez les gymnospermes où l'orientation des microtubules dans les trachéides suit étroitement l'orientation des microfibrilles de cellulose pendant la différenciation du xylème (Abe *et al.*, 1995 ; Funada *et al.*, 1997).



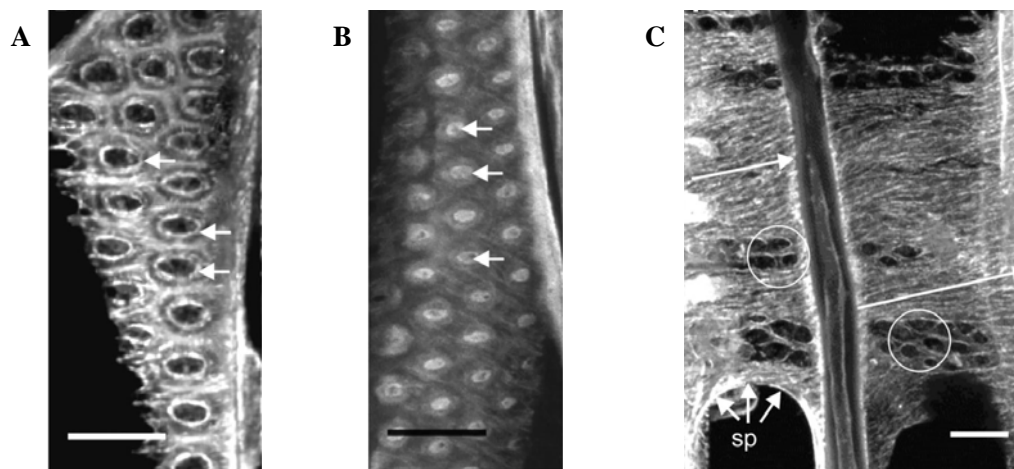
**Figure 13.** Organisation des microtubules dans les cellules cambiales et les fibres en différenciation chez le peuplier *P. tremula* × *P. tremuloides* (Chaffey *et al.*, 2002)

Les microtubules sont détectés par immunofluorescence indirecte de l' $\alpha$ -tubuline et visualisés en microscopie confocale laser (Échelle : 10  $\mu$ m).

(A) Les initiales cambiales en coupe tangentielle longitudinale. Les microtubules sont orientés au hasard dans les cellules fusiformes cambiales (f). Les microtubules sont aussi désordonnés dans les cellules de rayons (r) mais sont plus rares. (B) Coupe longitudinale tangentielle des fibres normales de xylème au stade précoce de différenciation correspondant à l'expansion cellulaire. Le réseau de microtubule hélicoïdal (flèche) se forme dans les fibres et les microtubules s'entrecroisent occasionnellement (cercle). (C) Coupe longitudinale radiale des fibres normales de xylème au stade tardif de différenciation correspondant à la synthèse de paroi secondaire. Les microtubules sont très denses et alignés parallèlement. Les flèches indiquent les différents angles d'orientation des microtubules dans les différentes fibres en cours de différenciation.

Bien que les mécanismes ne soient pas encore connus, il est maintenant communément admis que l'organisation des microtubules corticaux est étroitement liée à l'orientation des microfibrilles de cellulose et exerce une influence sur la synthèse de cellulose et la morphogenèse des cellules végétales (Burk & Ye, 2002 ; Gardiner *et al.*, 2003 ; Paredez *et al.*, 2006). De façon intéressante, certains gènes codants pour des tubulines, les unités de bases des microtubules, sont surreprésentés dans les banques d'ESTs de xylème en développement

par rapport à celle de zone cambiale (Hertzberg *et al.*, 2001 ; Déjardin *et al.*, 2004). Ces données confortent les observations montrant que dans les vaisseaux et fibres de xylème mettant en place leur paroi secondaire, le réseau de microtubules corticaux est plus dense que dans les cellules cambiales (Chaffey *et al.*, 1999 ; Chaffey *et al.*, 2002). Par ailleurs, les ponctuations simples des fibres se formant au début du stade de la formation de la paroi secondaire, sont dépourvues de microtubules. Du côté du phloème, les microtubules prennent aussi une organisation ordonnée et hélicoïdale lors de la différenciation de toutes les cellules axiales du phloème secondaire comme les tubes criblés, les cellules compagnes, et le parenchyme du phloème (Chaffey *et al.*, 2000b).



**Figure 14.** Organisation des microtubules dans les éléments de vaisseaux à différents stades de différenciation chez le peuplier *P. tremula* × *P. tremuloides* (Chaffey *et al.*, 2002)

Les microtubules sont détectés par immunofluorescence indirecte de l' $\alpha$ -tubuline et visualisés en microscopie confocale laser sur des coupes tangentielle radiales (Échelle : 20  $\mu$ m). (A) Développement des ponctuations aréolées dans les éléments de vaisseaux en cours de différenciation. Les cercles de microtubule indiqués par des flèches et entourant l'ouverture des ponctuations aréolées, prennent une apparence elliptique. (B) Stade final du développement des ponctuations aréolées dans les éléments de vaisseaux différenciés. Les ellipses de microtubules associées à l'ouverture des ponctuations aréolées ont un diamètre diminué. Entre les ponctuations aréolées, les microtubules corticaux sont arrangés hélicoïdalement et alignés parallèlement dans la cellule. (C) Les éléments de vaisseaux matures aux stades tardifs de différenciation présentant différentes organisations du réseau de microtubules, avec des cercles de microtubules à la périphérie de la perforation simple en développement (sp), et des ponctuations de contact (régions cerclées), et des microtubules transverses indiqués par des flèches dans le reste de la cellule.

Les éléments de vaisseaux à maturité présentent un réseau de microtubules corticaux dense, parallèle et orienté presque transversalement par rapport à l'axe de la cellule (Fig. 14). Toutefois, au niveau de ces éléments de vaisseaux, les ouvertures des ponctuations aréolées sont dépourvues de microtubules (Fig. 14A). Ces zones dépourvues de microtubules sont déjà présentes et prédéfinies au sein du réseau désordonné de microtubules des cellules cambiales fusiformes durant le stade d'élongation cellulaire (Chaffey *et al.*, 1997a ; Funada *et al.*, 1997 ;



Chaffey *et al.*, 1999 ; Chaffey *et al.*, 2002). Lors du développement des ponctuations aréolées des éléments de vaisseaux, un cercle de microtubule se forme à leur périphérie (Fig. 14A). Au fur et à mesure du développement des vaisseaux, le diamètre de la bordure des ponctuations aréolées diminue laissant une petite ouverture elliptique (Fig. 14B). Le cercle de microtubule marque l'épaississement localisé de la paroi cellulaire accompagnant le développement de ces bordures et décroît aussi en diamètre au fur et à mesure que les bordures des ponctuations aréolées se forment. Comme pour les ponctuations aréolées, les ponctuations de contact (non aréolées) entre les vaisseaux et les cellules des rayons ont une ouverture dépourvue de microtubules et lors de leurs formations développent sur leurs pourtours un cercle de microtubules corticaux (Fig. 14C). Mais contrairement aux ponctuations aréolées, celui-ci ne décroît pas en diamètre lors de la formation de ces ponctuations de contact. De façon intéressante, les ouvertures des ponctuations aréolées ne sont pas lignifiées alors qu'elles se développent durant l'épaississement de la paroi secondaire (Chaffey *et al.*, 1997b).

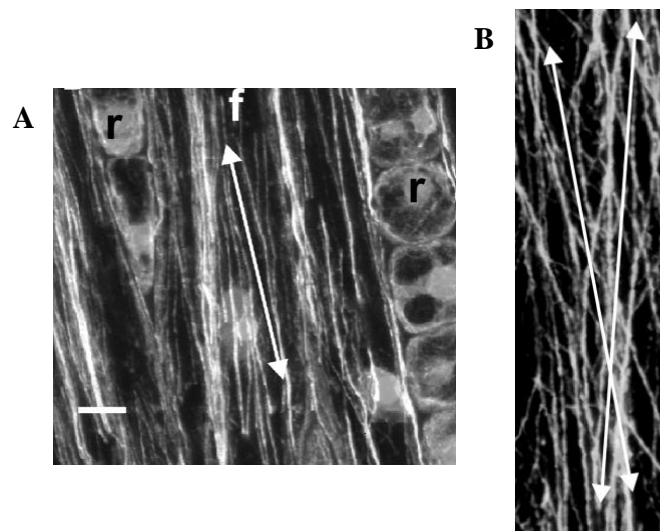
Les cellules de contact et parenchymateuse des rayons ont un réseau de microtubules désordonné pendant la division et l'élongation cellulaire. Plus tard, les microtubules prennent une orientation axiale dans les cellules de rayons et sont absents des ponctuations simples (Chaffey *et al.*, 2002). Les ponctuations de contact entre cellules de rayons portent aussi des cercles de microtubules corticaux et se développent de la même manière que celles entre les cellules de rayon et les éléments de vaisseaux. Quand les cellules de rayons développent de la paroi secondaire, les microtubules forment des réseaux denses orientés transversalement ou hélicoïdalement par rapport à l'axe radial de la cellule. Enfin, lorsque la synthèse de parois secondaire est terminée les microtubules sont alignés axialement dans toutes les cellules de rayons (Chaffey & Barlow, 2001 ; Chaffey *et al.*, 2002).

### 2.3.2. Les microfilaments d'actine

Chez le marronnier (*Aesculus hippocastanum*) et le peuplier (*P. tremula* × *P. tremuloides*), les microfilaments d'actine apparaissent agrégés dans des ensembles de paquets ou câbles d'actine parallèles au sein des initiales fusiformes (Chaffey *et al.*, 2002). Ces câbles d'actine sont alignés de façon relativement parallèle à l'axe longitudinal des initiales fusiformes, et apparaissent étendus entre les deux extrémités de la cellule (Fig. 15A). L'arrangement des microfilaments d'actine dans les initiales fusiformes cambiales est clairement distinct de celui des microtubules qui ont un arrangement désordonné. Cette

## Introduction Générale

orientation nettement axiale des microfilaments d'actine serait liée à la croissance intrusive aux deux extrémités des cellules fusiformes et permettrait le transport localisé des vésicules de sécrétion à ces endroits. Les câbles d'actine contribueraient au courant cytoplasmique nécessaire à la distribution des précurseurs dans ces longues cellules fortement vacuolisées. De plus, il a été montré par immunolocalisation que les microfilaments d'actine sont localisés le long de la plaque cellulaire des initiales fusiformes, et avec les microtubules, sont fortement associés au phragmoplaste (Chaffey & Barlow, 2002). Il a été suggéré que les microtubules et les microfilaments d'actine guident la migration du phragmoplaste (Mellerowicz *et al.*, 2001). Contrairement aux initiales fusiformes, les microfilaments d'actine n'ont pas une orientation axiale dans des initiales radiales, et sont comme les microtubules arrangés au hasard. De plus, comparées aux initiales fusiformes, les câbles d'actines sont plus rares dans les initiales radiales (Chaffey & Barlow, 2001 ; Chaffey *et al.*, 2002). Dans ces cellules cubiques et plus petites que les initiales fusiformes, le matériel nécessaire à la formation de la paroi est plus facilement acheminé par simple diffusion. Le besoin de courant cytoplasmique dans les initiales radiales est donc moins important que dans les initiales fusiformes, ce qui expliquerait la rareté des câbles d'actine dans les initiales radiales (Chaffey *et al.*, 2002).



**Figure 15.** Organisation des microfilaments d'actine dans les cellules cambiales et les fibres en différenciation chez le peuplier *P. tremula* × *P. tremuloides* (Chaffey *et al.*, 2002)

L'actine est détectée par immunofluorescence indirecte et la fluorescence émise est visualisée en microscopie confocale laser. (Échelle : 10 µm)

(A) Coupe tangentielle longitudinale de la zone cambiale. Les câbles d'actine marqués par des doubles flèches ont une orientation presque axiale dans les initiales fusiformes (f). Dans les initiales radiales (r), les câbles d'actine plus rares sont désordonnés.

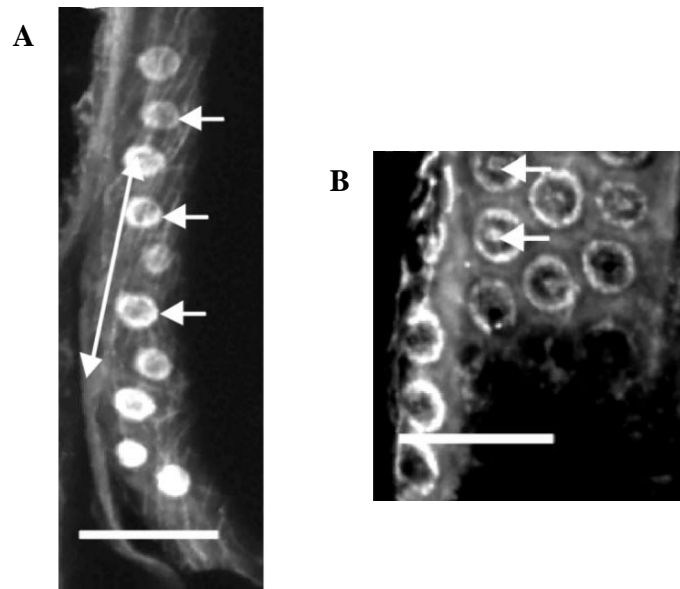
(B) Coupe longitudinale radiale de fibres normales de xylème. Les câbles de microfilaments d'actine indiqués par des doubles flèches, sont alignés axialement dans les fibres de xylème.

## Introduction Générale

Au cours du développement du xylème secondaire, l'orientation des microfilaments d'actine ne varie pas entre les initiales fusiformes et toutes les cellules secondaires qui en dérivent. Ainsi, les câbles de microfilaments d'actine gardent une orientation axiale et hélicoïdale dans les fibres de xylème en développement et mature (Fig. 15B), les vaisseaux (Fig. 15A), les cellules axiales parenchymateuses du xylème et du phloème, les cellules compagnes et les tubes criblés du phloème (Chaffey *et al.*, 2000a ; Chaffey *et al.*, 2002). Leur orientation longitudinale dans les fibres en élancement est en accord avec une croissance intrusive aux extrémités de ces cellules. Mais l'orientation similaire des microfilaments d'actine dans les fibres formant la paroi secondaire et ayant arrêté leur élancement implique d'autres fonctions non liées à la croissance intrusive de ces fibres. De plus, il y a aussi une remarquable augmentation de la fréquence du branchement latéral des microfilaments du réseau d'actine dans les éléments de vaisseaux et les fibres par rapport aux initiales fusiformes. Il apparaît que l'étendue du branchement des microfilaments d'actine est corrélée avec le degré de croissance en diamètre des cellules dérivant des initiales fusiformes, et serait lié au transport d'enzymes ou de précurseurs pariétaux vers des sites spécifiques sur les parois latérales comme les ponctuations simples (Chaffey *et al.*, 2000a ; Chaffey *et al.*, 2002). En outre, les fibres possèdent des ellipses de microfilaments d'actine au niveau des ponctuations simples. Ces aires sont dépourvues de microtubules et ne subissent pas d'épaississement de paroi, laissant penser que les microfilaments d'actine définissent une zone d'exclusion de microtubules. A la fin de la différenciation des fibres, les microfilaments d'actine et les microtubules deviennent plus clairsemés et désorganisés, et pendant la mort cellulaire programmée ils disparaissent de la cellule comme les autres organites cellulaires (Chaffey *et al.*, 2002).

Les éléments de vaisseaux à un stade de développement précoce présentent des réseaux circulaires de microfilaments d'actine associés avec la périphérie et l'ouverture des ponctuations aréolées et de contact (Fig. 16A) (Chaffey *et al.*, 2002). Lors du développement des ponctuations aréolées dans les éléments de vaisseaux plus âgés, le réseau circulaire de microfilaments d'actine est associé aux ouvertures des ponctuations, prend une forme elliptique, et a un diamètre plus réduit (Fig. 16B). Dans les éléments de vaisseaux matures, bien que les microfilaments d'actine maintiennent un réseau axial très branché, des microfilaments peuvent avoir une orientation transversale comme les microtubules. Ces microfilaments d'actines transverses pourraient ainsi coopérer avec les microtubules durant la phase d'épaississement de la paroi secondaire des éléments de vaisseaux. Contrairement aux initiales cambiales où les microfilaments d'actine et les microtubules ne sont pas en contact,

le degré de coopération est plus marqué entre ces deux éléments du cytosquelette dans les cellules du xylème qui en dérivent. Notamment, leur colocalisation au niveau des ponctuations aréolées et de contact suggère une implication des microtubules et des microfilaments d'actine dans le développement de ces ponctuations (Chaffey *et al.*, 2002).



**Figure 16.** Organisation des microfilaments d'actine dans les éléments de vaisseaux en différenciation chez le peuplier *P. tremula* × *P. tremuloides* (Chaffey *et al.*, 2002)

L'actine est détectée sur des coupes longitudinales radiales par immunofluorescence indirecte et la fluorescence émise est visualisée en microscopie confocale laser. (Échelle : 20 µm)

(A) Développement des ponctuations aréolées dans les éléments de vaisseaux en cours de différenciation (stade de développement similaire à celui de la figure 14A). Les câbles de microfilaments d'actine sont orientés axialement (flèche double) dans la cellule, et l'actine forme des cercles au niveau des sites de développement des ponctuations aréolées (flèches).

(B) Les ponctuations aréolées dans les éléments de vaisseaux à un stade plus tardif de développement (stade de développement similaire à celui de la figure 14B). L'actine forme des ellipses associées aux ouvertures en développement des ponctuations aréolées (flèches).

Les microfilaments d'actines sont plus fréquents dans les cellules de rayon que dans les initiales radiales. Ils sont alignés parallèlement à l'axe longitudinal de la cellule en élongation dès que la différenciation des cellules de rayon commence (Chaffey & Barlow, 2001 ; Chaffey *et al.*, 2002). Bien que la croissance de ces cellules soit plutôt de type diffus, cette orientation axiale des microfilaments faciliterait le transport de matériel pour l'élongation de ces cellules. Les ellipses de microfilaments sur les ponctuations simples des fibres et des cellules de rayon sont souvent attachées à des câbles d'actine se projetant dans le cytoplasme et qui guideraient le transport de vésicules vers ces ponctuations simples (Chaffey *et al.*, 2002). Lorsque la synthèse de la paroi secondaire est terminée, les microfilaments d'actines sont toujours orientés axialement, mais ils sont plus épais et empaquetés, et sont apparemment branchés entre eux. Les microfilaments d'actine et les microtubules sont

toujours présents dans les cellules de rayons matures. Les cellules de rayons constituent des voies symplasmiques importantes perméabilisant le tronc, du phloème jusqu'au le xylème, et permettant le transport des réserves et des produits de la photosynthèse tout au long de la vie de l'arbre. Il a été suggéré que le cytosquelette d'actine et de microtubule pourrait être impliqué dans le transport intracellulaire et intercellulaire de matériel au sein d'un continuum symplasmique tridimensionnel comprenant les cellules de rayons et les cellules axiales parenchymateuses (Chaffey & Barlow, 2001).

### *2.4. La biosynthèse de cellulose et les cellulose synthases chez les plantes*

#### 2.4.1. La biosynthèse de cellulose

Chez toutes les plantes inférieures et supérieures, la cellulose est un constituant essentiel des parois cellulaires primaires et secondaires. La teneur plutôt faible en cellulose dans la paroi primaire permet l'expansion et la croissance des cellules. Par contre, elle est le composé majeur des parois secondaires des cellules de bois représentant près de 50 % de leur poids sec. Elle confère ainsi à la paroi cellulaire secondaire une rigidité et un soutien mécanique. La cellulose est un homopolymère linéaire non branché de haut poids moléculaire formé d'unités  $\beta(1\rightarrow4)$  D-glucane (Brown *et al.*, 1996). Dans la paroi primaire, les chaînes de cellulose s'associent spontanément pour former des microfibrilles de cellulose. Les chaînes cellulosiques des microfibrilles sont alignées parallèlement et orientée de telle sorte que leurs extrémités réductrices se trouvent toutes du même côté. Dans la paroi secondaire, ces microfibrilles de cellulose s'associent entre elles par des liaisons hydrogène intra et interchaînes entre les résidus de glucose pour former des macrofibrilles (Delmer & Amor, 1995). Les chaînes de celluloses peuvent contenir plusieurs milliers de résidus glucose de 9800 dans les fibres de xylème de peuplier jusqu'à 15000 dans les fibres de coton. Le degré de polymérisation des résidus glucose varie aussi entre les parois primaire et secondaire (Brown *et al.*, 1996). La cellulose s'agrège naturellement dans une forme cristalline fortement ordonnée, avec aussi des formes amorphes moins ordonnées. La cellulose du bois est à 60-70 % cristalline et cette forme cristalline de cellulose possède une meilleure force de tension (Timell, 1969). Chez les plantes supérieures, la synthèse de cellulose se déroule au niveau de structures hexagonales en forme de rosettes et situées sur la membrane plasmique des cellules. Ces rosettes renferment six sous unités correspondant aux cellulose synthases (CesA), des

enzymes catalysant la formation de polymères de résidus glucoses liés en  $\beta(1\rightarrow4)$  à partir d'UDP-glucose. Chez le cotonnier et probablement toutes les plantes supérieures, l'initiation de la polymérisation de la chaîne de  $\beta(1\rightarrow4)$  D-glucane pourrait faire intervenir une glycosyltransférase (GhCesA-1) utilisant le sitostérol- $\beta$ -glucoside comme amorce (Peng *et al.*, 2002). Ces sitostérol- $\beta$ -glucosides sont synthétisés grâce à une  $\beta$ -sitostérol glucosyl transférase associée à la membrane plasmique. Les membranes de fibres de coton synthétisent des sitostérol-cellodextrines (des sitostérols comportant plusieurs résidus glucoses) à partir du sitostérol- $\beta$ -glucoside et de l'UDP-glucose dans des conditions favorisant la synthèse de cellulose. Ces sitostérol-cellodextrines doivent ensuite être clivés par une cellulase (ou endoglucanase), puis être transférés à la chaîne de  $\beta(1\rightarrow4)$  D-glucane en élongation (Peng *et al.*, 2002).

### 2.4.2. La famille et la structure des cellulose synthases (CesA) de plante

Le premier ADNc de plantes codant pour une cellulose synthase a été isolé chez le cotonnier à partir d'une banque d'ADNc de fibres de coton (Pear *et al.*, 1996). Ces fibres développent des parois secondaires constituées à 95 % de cellulose. Les cellulose synthases appartiennent à une petite famille multigénique et leurs relations phylogénétiques sont intéressantes en regard de la famille de protéines à domaine LIM que nous verrons plus tard. Les génomes du peuplier et d'*Arabidopsis* contiennent respectivement 18 et 10 gènes *CesA* (Tableau 4) (Richmond, 2000 ; Djerbi *et al.*, 2005 ; Tuskan *et al.*, 2006). De façon intéressante, une paire de gène (*PtCesA6-1* et *PtCesA6-2*) chez *P. trichocarpa* ne possède pas d'orthologue correspondant chez *A. thaliana* et sont très proches des cellulose synthases de graminées (riz et maïs) avec une identité en acides aminés supérieure à 80 % (Samuga & Joshi, 2004). Par ailleurs, le génome du peuplier possède 96 gènes liés à la synthèse de cellulose, alors qu'*Arabidopsis* n'en contient que 78. Par exemple, les gènes codant pour les gènes *KOR*, *FRA2*, *SuSy* (sucrose synthase) et *COBRA* sont représentés par des paires de gènes dupliqués chez le peuplier comparés aux homologues simple copie chez *Arabidopsis*. Il apparaît ainsi que les gènes associés à la synthèse de cellulose ont été dupliqués chez le peuplier et que les copies de ces gènes ont été conservées (Djerbi *et al.*, 2005 ; Tuskan *et al.*, 2006). Une analyse phylogénétique intéressante incluant trois nouveaux gènes *CesA* de pin (*Pinus taeda*) montre que chez les plantes à graines, les gènes *CesA* sont divisés en six groupes paralogues comprenant des gènes orthologues d'espèces différentes (Nairn &

## Introduction Générale

Haselkorn, 2005). Il a été ainsi suggéré que ces groupes de cellulose synthases paralogues ayant des fonctions distinctes ont évolué avant la divergence des lignées d'angiosperme et de gymnosperme. Les trois gènes *CesA* isolés chez *Pinus taeda* sont homologues aux cellulose synthases secondaires d'*Arabidopsis* et de peuplier, ont aussi un profil d'expression spécifique des tissus formant de la paroi secondaire, et pourraient avoir une fonction analogue aux orthologues isolés chez les angiospermes (Nairn & Haselkorn, 2005).

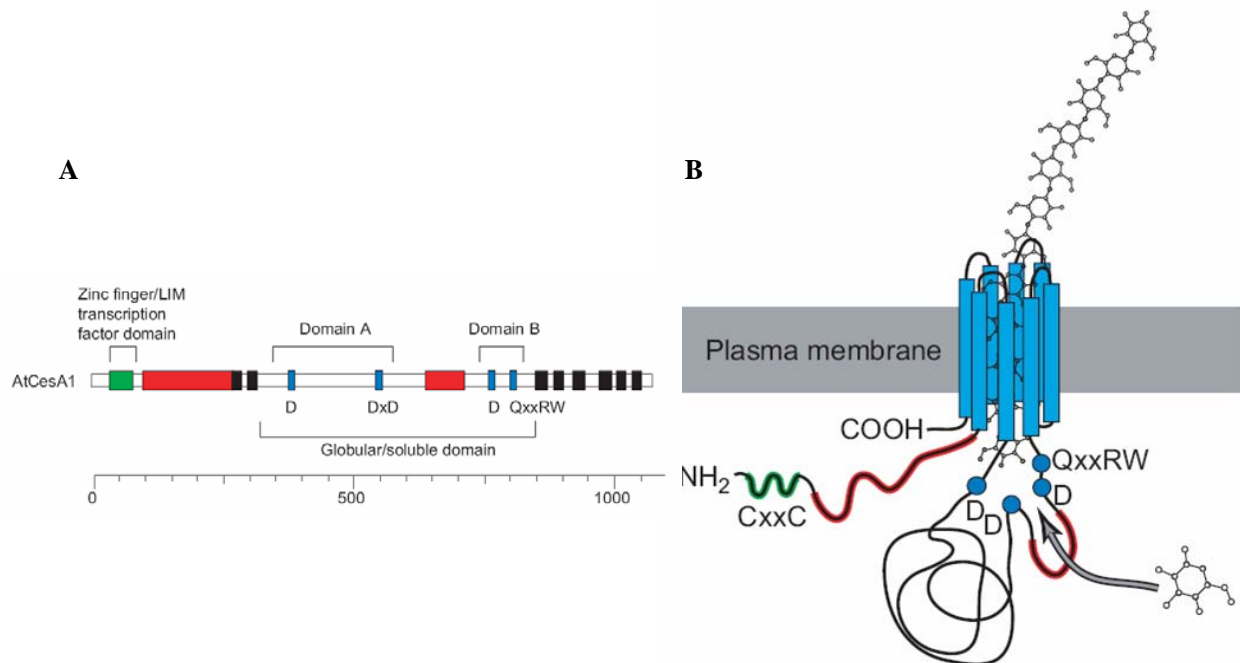
**Tableau 4.** Récapitulatif de la nomenclature des gènes codant pour des cellulose synthases chez le peuplier et *Arabidopsis*

Gènes <i>P. trichocarpa</i> (Djerbi, <i>et al.</i> , 2005)	Gènes <i>P. tremula</i> × <i>P. tremuloides</i> (Djerbi <i>et al.</i> , 2004)	Gènes <i>P. tremuloides</i> (Joshi <i>et al.</i> , 2004)	Gènes orthologues <i>A. thaliana</i> (Somerville, 2006)
<i>PtCesA1</i>	<i>PttCesA1</i>	<i>PtrCesA3</i>	<i>AtCesA4</i> ( <i>irx5</i> )
<i>PtCesA2-1</i> <i>PtCesA2-2</i>	<i>PttCesA2</i>		<i>AtCesA2</i> , <i>AtCesA5</i> , <i>AtCesA6</i> ( <i>irx2</i> , <i>prc</i> ) et <i>AtCesA9</i>
<i>PtCesA3-1</i> <i>PtCesA3-2</i>	<i>PttCesA3-1</i> <i>PttCesA3-2</i>	<i>PtrCesA/PtrCesA1</i>	<i>AtCesA8</i> ( <i>irx1</i> , <i>fra6</i> )
<i>PtCesA4-1</i> <i>PtCesA4-2</i>	<i>PttCesA4</i>	<i>PtrCesA7</i>	<i>AtCesA2</i> , <i>AtCesA5</i> , <i>AtCesA6</i> ( <i>irx2</i> , <i>prc</i> ) et <i>AtCesA9</i>
<i>PtCesA5-1</i> <i>PtCesA5-2</i>	<i>PttCesA5</i>	<i>PtrCesA5</i>	<i>AtCesA3</i> ( <i>irx1</i> , <i>cev1</i> , <i>eli</i> )
<i>PtCesA6-1</i> <i>PtCesA6-2</i>	<i>PttCesA6</i>	<i>PtrCesA6</i>	
<i>PtCesA7-1</i> <i>PtCesA7-2</i>	<i>PttCesA7</i>		<i>AtCesA3</i>
<i>PtCesA8-1</i> <i>PtCesA8-2</i> <i>PtCesA8-3</i>	<i>PttCesA8</i>	<i>PtrCesA4</i>	<i>AtCesA1</i> ( <i>rsw1</i> ) et <i>AtCesA10</i>
<i>PtCesA9-1</i> <i>PtCesA9-2</i>	<i>PttCesA9</i>	<i>PtrCesA2</i>	<i>AtCesA7</i> ( <i>irx3</i> , <i>fra5</i> )

Les gènes codant pour des cellulose synthases secondaires sont mis en évidence par un arrière fond gris. Les noms des différents mutants pour chaque gène *AtCesA* d'*Arabidopsis* sont indiqués entre parenthèses.

Les cellulose synthases sont des protéines membranaires comportant des motifs intervenant dans son activité enzymatique et dans les interactions avec les autres protéines *CesA* pour former des hexamères (Fig. 17). Il est intéressant de noter la présence à l'extrémité N-terminale des protéines *CesA* d'un domaine portant certaines ressemblances avec les doigts de zinc des domaines RING ou LIM que nous verrons plus tard. Au sein de ce domaine le motif CxxC est très conservé et la séquence consensus suivante a été déterminée [CX<sub>2</sub>CX<sub>12</sub>FXACX<sub>2</sub>CX<sub>2</sub>PXCX<sub>2</sub>CX-EX<sub>5</sub>GX<sub>3</sub>CX<sub>2</sub>C] (Richmond, 2000). Une analyse structurale fine des domaines RING chez les plantes a exclu ce type de domaine à doigt de zinc de la famille de protéine à domaine RING (Kosarev *et al.*, 2002). Pourtant, il y a de forts arguments supportant un rôle du domaine à doigt de zinc des protéines *CesA* dans les

interactions protéine/protéine. Dans le cas de la protéine GhCesA1 de cotonnier, ce domaine peut lier deux atomes de zinc uniquement en condition réductrice et des expériences de double hybride chez la levure ont montré que les portions N-terminales des protéines GhCesA1 et GhCesA2 peuvent former des homo et hétérodimères. Cette interaction se ferait par des ponts disulfures et en condition oxydante (Kurek *et al.*, 2002). En plus de ce domaine liant le zinc, les celluloses synthases contiennent deux régions variables, huit domaines transmembranaires prédits, et entre ces domaines transmembranaires, un domaine globulaire soluble. Ce domaine globulaire contient deux domaines A et B possédant respectivement les motifs [D suivi de DXD] et [QXXRW] (Fig. 17A). Le domaine A pourrait être le domaine fixant l'UDP-glucose et le domaine B serait le domaine catalytique (Richmond, 2000).



**Figure 17.** Structure des cellulose synthases de plantes (Richmond, 2000).

**(A)** Représentation schématique des différents domaines d'une cellulose synthase en prenant la protéine AtCesA1 d'*Arabidopsis* comme exemple.

**(B)** Structure tridimensionnelle hypothétique des protéines CesA. Les hélices transmembranaires des protéines CesA forment un pore dans la membrane plasmique à travers lequel la chaîne de glycanes en croissance passerait. Les boîtes ou régions colorées en bleu, rouge, vert et noir correspondent respectivement aux résidus conservés des domaines A et B, aux régions variables, au domaine à doigt de zinc (Zinc finger) contenant le motif CxxC, et aux domaines transmembranaires.

### 2.4.3. Les cellulose synthases d'*Arabidopsis thaliana*

Les différentes sous unités de cellulose synthase ont été largement étudiées chez *Arabidopsis* grâce aux mutants « *irregular xylem* » (*irx*) et « *isoxaben-resistant* » (*ixr*) qui ont permis des avancées considérables dans la compréhension des mécanismes de biosynthèse de



## Introduction Générale

la cellulose (Somerville, 2006). Le premier mutant qui a permis de prouver que les cellulose synthases de plantes sont responsables de la synthèse de cellulose est le mutant « *radial swelling* » (*rsw1*) correspondant au gène *AtCesA1* (Arioli *et al.*, 1998). En effet, chez ce mutant, il a été observé une diminution du nombre de rosettes et de la quantité de cellulose cristalline. Il a été proposé que ce gène soit nécessaire à la synthèse de paroi cellulaire primaire. De même, l'analyse des mutants *ixr1* et *ixr2* touchant respectivement les gènes *AtCes3* et *AtCes6*, montre que ces deux gènes sont également impliqués dans la biosynthèse de la cellulose de la paroi primaire. Ces trois gènes *AtCesA1*, *AtCesA3* et *AtCesA6*, et probablement aussi les gènes *AtCesA2* et *AtCesA5*, sont donc considérés comme des cellulose synthases fonctionnant dans l'élaboration de la paroi primaire (Somerville, 2006). Il a récemment été identifié un mutant knock-out pour le gène *AtCesA2*, et chez ce mutant la quantité de cellulose diminue aussi fortement, mais uniquement dans les parties aériennes (Chu *et al.*, 2007). De plus, les plantes mutantes *atcesa2* ont une croissance fortement réduite, des cellules plus courtes, et une altération de l'orientation des microtubules.

Alors que la fonction des gènes *AtCesA9* et *AtCesA10* dans la biosynthèse de cellulose n'est pas encore connue, les trois autres cellulose synthases *AtCesA4*, *AtCesA7* et *AtCesA8*, dont les mutants respectifs sont *irx5*, *irx1* et *irx3*, ont un rôle dans la synthèse de cellulose de la paroi cellulaire secondaire (Somerville, 2006). Les mutants *irx5*, *irx1/fra6* et *irx3/fra5* touchent les mêmes types cellulaires (le xylème en développement), et présentent tous le même phénotype de xylème collapsé dû à une diminution de la quantité de cellulose dans la paroi secondaire (Taylor *et al.*, 1999 ; Taylor *et al.*, 2000 ; Taylor *et al.*, 2003). Ces trois gènes sont exactement co-exprimés dans les mêmes cellules, et il a été montré que les protéines *AtCesA7* et *8* ne s'associent plus en l'absence de la protéine *AtCesA4* (Taylor *et al.*, 2003). Ces trois protéines sont ainsi des composants essentiels au complexe de synthèse de la cellulose et l'absence de l'une des trois protéines provoque la rétention des deux autres protéines dans le cytoplasme de la cellule (Gardiner *et al.*, 2003). Enfin, les protéines *AtCesA4*, *AtCesA7* et *AtCesA8* co-localisent avec le réseau de microtubules cortical dans le xylème en développement. De plus, l'utilisation d'Oryzaline montre que les microtubules sont nécessaires à la localisation de la protéine *AtCesA8*. Mais contrairement au mutant *atcesa2*, chez les mutants *irx1*, *3* ou *5*, où aucune protéine *CesA* secondaire n'est localisée sur la membrane plasmique, l'organisation des microtubules dans le xylème est normale. Ainsi, dans la paroi secondaire, la présence du complexe de cellulose synthase n'est probablement pas nécessaire à la localisation correcte des microtubules. De façon intéressante, les deux mutants « *fragile fiber* » *fra5* et *fra6* montrent une réduction de la solidité mécanique des

tiges, et chez le mutant *fra5*, l'épaisseur des parois et la teneur en cellulose diminuent fortement (Zhong *et al.*, 2003). Ces analyses montrent l'importance du rôle de la cellulose dans la résistance des tiges à la force de cassure.

### 2.4.4. Les cellulose synthases de peuplier

Le premier ADNc codant pour une cellulose synthase de peuplier, PtCesA renommée plus tard PtrCesA1, a été isolé en 2000 chez *P. tremuloides* (Wu *et al.*, 2000). Depuis, beaucoup de données ont été accumulées sur l'expression des gènes codant pour des celluloses synthases chez le peuplier (Joshi, 2003 ; Djerbi *et al.*, 2004 ; Joshi *et al.*, 2004). Les gènes *PtrCesA1*, *PtrCesA2* et *PtrCesA3*, ont été appelé Cesa secondaires et sont respectivement homologues aux gènes *AtCesA8*, *AtCesA4* et *AtCesA7* d'*Arabidopsis* préférentiellement exprimés dans les tissus vasculaires secondaires (Joshi *et al.*, 2004 ; Somerville, 2006). Ces trois gènes (*PtrCesA1* à 3) sont fortement exprimés et de façon coordonnée dans les mêmes tissus : les xylème et phloème secondaires en développement (Samuga & Joshi, 2002 ; Joshi *et al.*, 2004 ; Kalluri & Joshi, 2004 ; Bhandari *et al.*, 2006). Dans les portions de tiges ayant une expansion radiale, les analyses en hybridation *in situ* montrent que les gènes *PtCesA1* et *PtCesA3* sont uniquement exprimés dans la première rangée de cellules du xylème secondaire (Wu *et al.*, 2000 ; Kalluri & Joshi, 2004). Alors que le gène *PtrCesA1* est aussi exprimé dans les cellules de xylème primaire, l'expression du gène *PtrCesA3* est restreinte aux cellules formant de la paroi secondaire (Kalluri & Joshi, 2004). Par ailleurs, il a été montré chez le tabac que le promoteur du gène *PtrCesA1* dirige l'expression du gène rapporteur *GUS* spécifiquement dans le xylème quels que soient les organes et en condition normale de croissance (Wu *et al.*, 2000). De la même manière, des analyses d'expression microarray et RT-PCR quantitative montrent que l'expression des gènes *PttCesA1*, *PttCesA3-2* et *PttCesA9-1* chez *P. tremula* × *P. tremuloides* est spécifique du xylème et est consistante avec l'expression des gènes homologues chez *P. tremuloides* (*PtrCesA3*, *PtrCesA1* et *PtrCesA2* respectivement ; Tableau 4) (Djerbi *et al.*, 2004). De plus, l'expression du gène dupliqué *PttCesA3-1* est similaire à celle du gène *PttCesA3-2* suggérant une redondance de fonction dans le xylème en développement. Enfin, de façon surprenante, le gène *PtrCesA5*, l'orthologue du gène *AtCesA3* codant pour une cellulose synthase primaire, est plus fortement exprimé dans le xylème en développement que dans les tissus à croissance primaire (Kalluri & Joshi, 2003).

## Introduction Générale

Les gènes *PtrCesA4*, *PtrCesA6* (*PttCesA6*), *PtrCesA7* (*PttCesA4*) et *PttCesA2* présentent un profil d'expression qui serait plutôt associé au développement de la paroi cellulaire primaire (Djerbi *et al.*, 2004 ; Joshi *et al.*, 2004 ; Kalluri & Joshi, 2004 ; Samuga & Joshi, 2004). En effet, dans la tige et la racine mature, l'expression du gène *PtrCesA4* est localisée dans des cellules formant de la paroi primaire comme le cambium. Notons que le gène *PtrCesA4* est aussi exprimé dans les cellules de xylème synthétisant de la paroi secondaire au niveau des pétioles, jeunes tiges et cellules de rayon de tiges matures (Kalluri & Joshi, 2004). Les gènes *PtrCesA6* et 7 montrent un faible niveau d'expression dans une grande variété de tissus (Samuga & Joshi, 2004). En hybridation *in situ*, les transcrits *PtrCesA6* sont observés dans les cellules épidermiques et dans les tissus vasculaires de la tige où l'expression du gène est restreinte aux cellules de rayons et à la phase précoce du développement du xylème. Le profil d'expression du gène *PtrCesA7* est différent avec une forte expression au niveau de l'apex des tiges, du cylindre central de la racine, et dans le cambium, le phloème et le cortex de la tige (Samuga & Joshi, 2004). Ainsi, il existerait chez toutes les plantes supérieures différentes isoformes de protéines CesA impliquées soit dans la formation de la paroi primaire, soit dans la formation de la paroi secondaire. Toutefois ces fonctions hypothétiques doivent être confirmées expérimentalement chez le peuplier par des analyses de plantes transgéniques modifiées dans l'expression de ces gènes.

### 2.5. La lignine, un composé majeur des plantes et important lors de la formation du bois

La lignine est un hétéropolymère hydrophobe du métabolisme secondaire provenant de la voie de biosynthèse des phénylpropanoïdes. Elle est unique aux plantes et c'est comme la cellulose un composé majeur des parois secondaires. La lignine joue un rôle essentiel dans le soutien mécanique de l'arbre et dans le transport d'eau en imperméabilisant la paroi des vaisseaux. Elle permet une résistance des parois cellulaires à la pression liée au transport de sève brute. D'autre part, la lignine est également essentielle dans la réponse à la blessure et à l'infection par les pathogènes. C'est en effet un composé difficilement dégradable par les microorganismes et elle constitue ainsi une barrière physique protégeant les tissus. Dans certain cas comme la formation des fibres de bois, la lignification des cellules est suivie d'une mort cellulaire programmée. La biosynthèse de la lignine est donc un mécanisme d'adaptation de la plante au soutien de son propre poids, au transport d'eau et à la défense contre les pathogènes (Rogers & Campbell, 2004).

### 2.5.1. Lignification de la paroi cellulaire secondaire

Dans les cellules du xylème secondaire en développement, la lignification est initiée au niveau des angles cellulaires et progresse le long de la lamelle moyenne dès que la couche S1 est entièrement synthétisée et que la formation de la couche S2 est engagée (Donaldson, 2001 ; Grünwald *et al.*, 2002). Ces sites sont appelés sites de nucléation et permettent la croissance du polymère de lignine. Le dépôt de la lignine est décalé dans le temps par rapport à la formation des différentes couches de la paroi secondaire et est précédé par la déposition des hydrates de carbone. La couche S1 se lignifie quand la matrice de polysaccharides de la couche S2 est entièrement formée. Enfin, lorsque la couche S3 est élaborée, la lignification devient plus intense par l'incorporation de lignine dans l'épaisse couche S2. Généralement, la concentration en lignine est plus élevée dans la lamelle moyenne et les angles cellulaires que dans la couche S2 de la paroi secondaire, mais cette dernière occupant un volume plus important dans le bois, la paroi secondaire a le contenu en lignine le plus élevé. La nature chimique de la matrice de polysaccharides et l'orientation des microfibrilles de cellulose influencent le dépôt de lignine. Dans la lamelle moyenne et la paroi primaire, la lignine forme des structures sphériques, alors que dans la paroi secondaire, la lignine forme des lamelles qui suivent l'orientation des microfibrilles (Donaldson, 2001). Durant sa déposition, la lignine pourrait former des liens chimiques avec les hémicelluloses de la paroi, éliminant graduellement l'eau pour former un environnement hydrophobe.

### 2.5.2. La composition des polymères de lignine

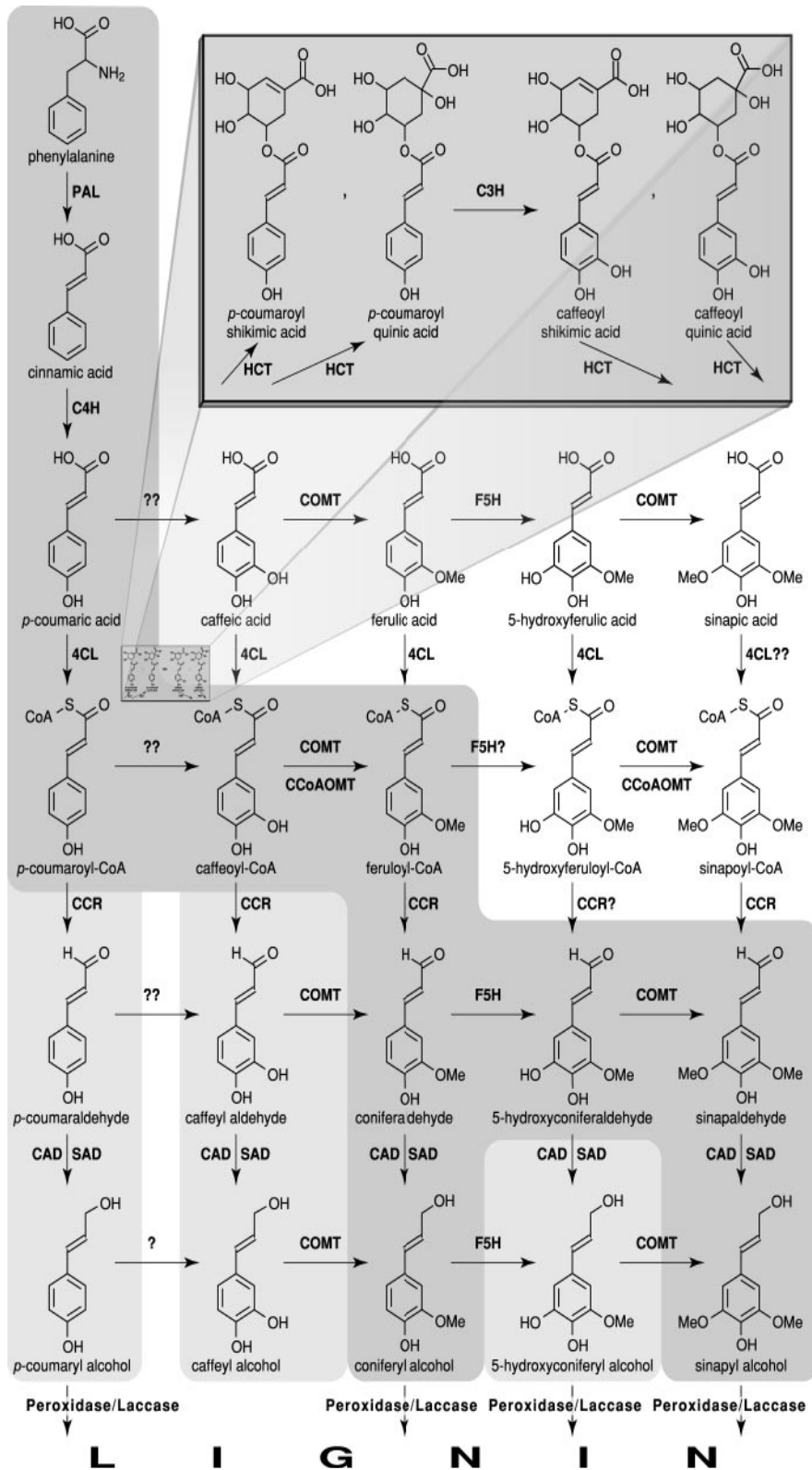
Ce polymère est classiquement composé de trois sous-unités *p*-hydroxyphényle (H), guaiacyle (G), et syringyle (S) dérivant respectivement de trois alcools cinnamiques (monolignols) : les alcools *p*-coumarique, coniferylque et synapilique. Ces sous-unités H, G et S diffèrent par le degré de méthylation aux positions 3' et/ou 5' du cycle aromatique (Fig. 18). Les réactions de méthylation et d'hydroxylation sont donc importantes dans la détermination de la composition de la lignine. D'autres sous-unités peuvent être incorporées à ce polymère comme des aldéhydes cinnamiques. De plus, des traces d'autres monomères provenant apparemment de la biosynthèse incomplète de monolignols sont souvent observées chez les plantes (Whetten *et al.*, 1998 ; Boerjan *et al.*, 2003). La lignine a une composition très variable en fonction des espèces, des contraintes environnementales et des types

cellulaires. Par exemple, chez les conifères, la lignine est composée majoritairement de sous unité G avec une faible teneur en unité S et H. Chez les angiospermes, la lignine est principalement constituée à la fois des deux sous unité G et S, et de quelques traces d'unité H. La lignine des plantes herbacées est composée de ces trois types de monomères et comporte une quantité importante de sous unité H (Boerjan *et al.*, 2003). La composition en lignine varie aussi en fonction de l'âge des tissus et du développement de la plante. Chez *Arabidopsis*, les tissus jeunes proches de l'apex contiennent moins d'unité S que les tissus matures à la base de la tige (Meyer *et al.*, 1998). Le contenu et la composition en lignine diffèrent également en fonction du type cellulaire, et même au sein de la paroi cellulaire (Boerjan *et al.*, 2003). Ainsi la lignification est plus précoce dans les éléments de vaisseaux et les cellules des rayons que dans les fibres. Les éléments de vaisseaux contiennent généralement plus de lignine que les fibres. De plus, les éléments de vaisseaux, les cellules de parenchyme des rayons et les fibres n'ont pas la même composition en lignine. En effet, la lignine dans les vaisseaux est généralement enrichie en unités G, alors que la lignine des cellules de rayons et des fibres contient habituellement plus d'unité S. Enfin, le type d'unité incorporé dans les polymères de lignine varie en fonction de la différenciation des cellules du xylème. Chez les angiospermes, les unités H sont habituellement déposées en premier dans la lamelle moyenne au stade précoce de différenciation, tandis que durant des stades plus tardifs du développement du xylème, ce sont plutôt des unités de type G et S qui sont incorporées dans la paroi secondaire. Cette hétérogénéité est probablement due à des différences dans la régulation spatiotemporelle de l'expression de certains gènes codant pour les enzymes de la voie de biosynthèse des lignines, et à des différences dans la spécificité de substrat de ces enzymes (Boerjan *et al.*, 2003).

### 2.5.3. La voie de biosynthèse de la lignine

Comportant un grand nombre de voies enzymatiques complexes telles que les voies des shikimates, phénylpropanoïdes et monolignols, la voie de biosynthèse des lignines commence à être bien connue. Elle débute par une désamination de la phénylalanine par la phénylalanine ammonia-lyase (PAL) pour donner l'acide cinnamique. A partir de ce dernier, des hydroxylations, des transferts de méthyles et de coenzymeA, et des réductions sur le noyau aromatique amènent à la formation des aldéhydes cinnamiques (Fig. 18).

## Introduction Générale



**Figure 18.** Voies de biosynthèse des phénylpropanoïdes et des monolignols (Boerjan *et al.*, 2003)

La voie la plus favorable pour la production des monolignols chez les angiospermes est encadrée en gris foncé. La voie en gris clair peut aussi être utilisée en fonction des espèces et des conditions physiologiques, et la voie non grisée ne joue pas de rôles significatifs dans la biosynthèse des monolignols.

## Introduction Générale

Ceux-ci sont réduits en alcool par une enzyme la cinnamyl alcool déshydrogénase (CAD) qui est la dernière enzyme de la voie de biosynthèse des monolignols. Les monolignols sont ensuite véhiculés probablement sous la forme de glucoside de monolignol vers la paroi où ils sont polymérisés après avoir été oxydés par des peroxydases, laccases, polyphénol oxydases ou alcool coniférique oxydases (Boerjan *et al.*, 2003). Toutefois, le rôle de ces enzymes dans la polymérisation de la lignine n'a pas encore été prouvé. De plus, les analyses fonctionnelles de ces gènes sont difficiles car les laccases et les peroxydases sont codées par des familles multigéniques, surtout les peroxydases avec au moins 73 gènes chez *Arabidopsis*. Bien que certains gènes codant pour des laccases chez le peuplier et le pin sont fortement voire exclusivement exprimés dans le xylème des tiges (Boerjan *et al.*, 2003), la dérégulation de l'expression de certaines laccases chez le peuplier ne semble pas affecter le contenu et la composition en lignine (Ranocha *et al.*, 2002). Cependant, la suppression de l'expression du gène *lac3* conduit à une altération de la paroi des fibres de xylème de peuplier. Alors que les laccases utilisent l'oxygène pour oxyder les monolignols, les peroxydases utilisent le peroxyde d'hydrogène (H<sub>2</sub>O<sub>2</sub>) et sont capables de catalyser l'oxydation des alcools conifériques. De plus, beaucoup de gènes codant pour des peroxydases anioniques et cationiques sont exprimés dans les cellules lignifiées du xylème, et la surexpression de certaines peroxydases peut générer une synthèse ectopique de lignine dans la plante (Boerjan *et al.*, 2003). Enfin, la polymérisation de la lignine pourrait être d'abord ordonnée et guidée par des protéines dirigeantes lors de la dimérisation des monolignols en lignanes, puis par la matrice de carbohydrates et l'orientation des microfibrilles de cellulose (Burlat *et al.*, 2001).

### 2.5.4. Les gènes codant pour les enzymes de la voie de biosynthèse des lignines

Les enzymes de la voie de biosynthèse des lignines font en général partie de petites familles multigéniques. Le génome du peuplier contient au moins 46 modèles de gène codant pour des enzymes de la voie de biosynthèse des monolignols (Tableau 5) (Tsai *et al.*, 2006 ; Tuskan *et al.*, 2006). La taille des familles de gènes codant pour ces enzymes chez *Populus trichocarpa* est généralement plus large que chez *Arabidopsis thaliana*. En effet, cette dernière comprend en tout 24 gènes intervenant dans la voie de biosynthèse des lignines (Raes *et al.*, 2003 ; Tsai *et al.*, 2006). Il existe donc pour chacune des familles d'enzymes impliquées dans la voie de monolignols plusieurs isoformes qui peuvent avoir des spécificités

## Introduction Générale

de substrat et des cinétiques enzymatiques différentes. De plus, les gènes codant ces isoformes sont en général exprimés différemment durant le développement de la plante, en fonction du type cellulaire et sous l'effet de facteurs environnementaux (Boerjan *et al.*, 2003). Cette grande flexibilité de la voie métabolique des phénylpropanoïdes est également illustrée par l'existence de boucles de rétrocontrôle régulant l'expression ou l'activité de certaines enzymes par l'intermédiaire de métabolites de la voie, comme par exemple la phénylalanine (Anterola *et al.*, 2002). Bien que les connaissances accumulées sur les enzymes de la voie de biosynthèse des lignines chez les plantes soient passionnantes (Boerjan *et al.*, 2003), dans la suite de cette partie, je ne traiterai que des familles d'enzymes PAL, 4CL et CAD, dont certains gènes sont potentiellement régulés par les facteurs de transcription LIM.

**Tableau 5.** Liste récapitulative du nombre de gènes codant pour des enzymes de la voie de biosynthèse des lignines chez *Arabidopsis thaliana* et *Populus trichocarpa*. (Raes *et al.*, 2003 ; Tsai *et al.*, 2006 ; Tuskan *et al.*, 2006)

Famille de gènes	Nombre de gènes chez <i>P. trichocarpa</i>	Nombre de gènes chez <i>A. thaliana</i>
phénylalanine ammonia-lyase (PAL)	5	4
trans-cinnamate 4-monooxygénase (C4H)	2 (1)	1
4-coumarate:CoA ligase (4CL)	5	4 (9)
p-coumaroyl shikimate/coumaroyl 3-hydroxylase (C3H)	3	1 (2)
coniférylaldehyde/férule 5-hydroxylase (F5H)	2 (1)	1 (1)
cinnamoyl CoA réductase (CCR)	7	2 (5)
alcool cinnamylique déshydrogénase (CAD)	1 (15)	2 (7)
trans-cafféoyl-CoA 3-O-méthyltransférase (CCoAOMT)	6	7
acide cafféique 3-O-méthyltransférase (COMT)	9	1 (13)
hydroxycinnamoyl CoA:shikimate/quinat hydroxycinnamoyltransférase (HCT)	6 (1)	1
Total	46 (18)	24 (37)

Les chiffres entre parenthèses indiquent le nombre de gènes similaires qualifiés comme « *-like* » pour chaque famille. Il est fort possible que le nombre de gènes « *-like* » chez *P. trichocarpa* soit sous estimé.

### 2.5.5. Les phénylalanine ammonia lyases (PAL)

La phénylalanine ammonia lyase (PAL) est la première enzyme de la voie générale des phénylpropanoïdes et catalyse la déamination de la phénylalanine en acide trans-cinnamique. Elle est impliquée dans la voie de biosynthèse des lignines, et est également importante dans d'autres voies enzymatiques telle que la voie de biosynthèse des flavonoïdes ou des anthocyanines. Les gènes codant pour des PAL sont exprimés dans une grande diversité de tissus et répondent à un grand nombre de signaux environnementaux et développementaux (Rogers & Campbell, 2004). La PAL est importante dans l'épiderme pour la protection contre



les UV, dans les pétales pour la production de pigments, dans la réponse aux pathogènes ou à la blessure et dans la biosynthèse de monolignols dans le xylème. Chez *Arabidopsis*, la famille de gène *PAL* est composée de quatre membres *PAL1* à *4* répartis en deux groupes comprenant respectivement *PAL1* et *2*, et *PAL3* et *4* (Raes *et al.*, 2003). L'expression des gènes *PAL1*, *PAL2* et *PAL4* augmente en fonction du stade de développement de la tige, alors que le gène *PAL3* est fortement exprimé dans la feuille. Les promoteurs des gènes *PAL1* et *PAL2* contiennent des éléments AC très conservés conférant une spécificité d'expression dans le xylème. Ces deux gènes sont donc des candidats potentiels impliqués dans la lignification des tissus vasculaires de la tige d'*Arabidopsis* (Ohl *et al.*, 1990 ; Raes *et al.*, 2003). De façon surprenante, le double mutant *pallpal2* ne présente que des altérations phénotypiques limitées (Rohde *et al.*, 2004). Toutefois de profonds changements transcriptomique et métabolomique sont observés chez ce double mutant. Il est notamment observé une altération de l'expression des gènes codant pour des enzymes de la biosynthèse des monolignols dont les gènes *PAL3* et *4*, une suraccumulation de phénylalanine, une teneur altérée des autres acides aminés, une diminution significative du contenu en lignine, une augmentation du rapport en unité S/G, et aussi une diminution de trois glycosides de flavonols majeurs (Rohde *et al.*, 2004). De même, la sous régulation de la PAL chez le tabac réduit largement le contenu en lignine de ces plantes affectant surtout le contenu en unités G, et provoque aussi une diminution du niveau d'acide chlorogénique (Elkind *et al.*, 1990 ; Bate *et al.*, 1994). La PAL devient limitante quand son activité est réduite en dessous d'un seuil de 20 à 25 % chez les tabacs transgéniques. Chez les espèces de peuplier, il existe cinq isoformes *PAL1* à *5* qui forment deux groupes phylogénétiques distincts. Les gènes *PAL4* et *5* montrent une très forte homologie de séquence et codent pour des protéines ayant 98,5 % d'identité en acides aminés (Tsai *et al.*, 2006 ; Tuskan *et al.*, 2006). Les gènes *PAL1* et *3* du groupe 1 sont exprimés dans les tissus riches en tanins condensés comme les feuilles, et les gènes *PAL2* et *PAL4/5* du groupe 2 sont exprimés dans les tissus lignifiés comme le xylème des tiges et les racines (Kao *et al.*, 2002 ; Tsai *et al.*, 2006).

### 2.5.6. Les 4-coumarate:CoA ligases (4CL)

Les 4-coumarate:CoA ligases (4CL) catalysent la formation d'esters de coenzyme A à partir d'une grande variété de substrats : les acides p-coumarique, caféique, férulique, 5-hydroxyférulique et sinapique. La famille 4CL chez les plantes comporte un grand nombre de

## Introduction Générale

gènes avec des protéines 4CL divisées en deux classes, et des 4CL-like formant un groupe monophylétique distinct. Chez *Arabidopsis*, les quatre vraies At4CL1 à 4 définissent la classe I (At4CL1, 2 et 4) et la classe II (At4CL3), et les At4CL-like sont au nombre de neuf (At4CL-like1 à 9) (Ehlting *et al.*, 1999 ; Raes *et al.*, 2003). Les gènes *At4CL1* et 2 sont fortement exprimés dans les tiges et leur expression augmente pendant les stades tardifs du développement de la tige, ainsi que dans les feuilles après une blessure. Ces deux gènes *At4CL1* et 2 pourraient être spécifiquement impliqués dans la lignification et contiennent des éléments AC dans leurs promoteurs. A l'inverse, le gène *At4CL3* de la classe II, plus spécifiquement exprimé dans les feuilles, les fleurs et les siliques, et avec une expression fortement induite par des radiations UV, pourrait participer à la voie de biosynthèse des flavonoïdes (Elkind *et al.*, 1990 ; Raes *et al.*, 2003). Le gène *At4CL4* est faiblement exprimé dans les feuilles et les tiges et contrairement aux protéines 4CL1, 2, et 3, la protéine 4CL4 serait capable d'utiliser l'acide sinapique comme substrat (Ehlting *et al.*, 1999 ; Raes *et al.*, 2003 ; Hamberger & Hahlbrock, 2004). La famille 4CL chez les espèces de peuplier comprend cinq membres avec des paires de gènes dupliqués (*4CL1/5* et *4CL3/4*) (Tsai *et al.*, 2006). Chez le peuplier tremble (*P. tremuloides*), le gène *Pt4CL1* de la classe I est orthologue aux gènes *At4CL1*, 2 et 4, et le gène *Pt4CL2* de la classe II est orthologue au gène *At4CL3*. De la même manière que chez *Arabidopsis*, le gène *Pt4CL1* est exprimé dans les cellules lignifiées du xylème. La protéine Pt4CL1 utilise sélectivement l'acide caféique comme substrat et serait plutôt impliqué dans la lignification. Au contraire, la protéine Pt4CL2 produit préférentiellement du p-coumaroyl-CoA et serait impliquée dans la formation d'autres composés phénoliques comme les flavonoïdes dans les cellules épidermiques des feuilles (Hu *et al.*, 1998 ; Harding *et al.*, 2002). Le gène *Pt4CL5* de peuplier est exprimé de la même manière que le gène *Pt4CL1* avec une forte expression dans le xylème (Tsai *et al.*, 2006). Les gènes *4CL3* et 4 de la classe I sont fortement exprimés dans les vieilles et les jeunes feuilles respectivement. Les protéines correspondantes présentent un profil d'utilisation de substrat similaire avec une affinité uniquement pour les acides p-coumarique, férulique et caféique (Allina *et al.*, 1998 ; Tsai *et al.*, 2006). La sous régulation de la 4CL chez le tabac, *Arabidopsis* et le peuplier conduit à une réduction de la teneur en lignine, et à un plus fort taux d'acides hydroxycinnamique liés à la paroi cellulaire (Boerjan *et al.*, 2003). L'effet sur la composition en unité S et G de la lignine est variable en fonction des espèces considérées. Une explication possible est que différentes isoformes ou de multiples isoformes 4CL ayant des spécificités de substrat et de fonction distinctes peuvent être différenciellement touchées dans ces plantes transgéniques. Une diminution de l'activité 4CL chez des plantes

transgéniques d'*Arabidopsis* conduit à une diminution du ratio G/S. Chez ces plantes, la réduction forte du contenu en lignine n'est généralement pas associée avec des anomalies de croissance (Lee *et al.*, 1997). Dans le cas de la dérégulation de la Pt4CL1 chez le peuplier tremble où la réduction de la lignine est de 45 %, la croissance des peupliers est améliorée, les parois cellulaires gardent leur intégrité, et la diminution du contenu en lignine est compensée par l'incorporation d'autres constituants de la paroi notamment la cellulose dont le contenu augmente de 15 % (Hu *et al.*, 1999).

### 2.5.7. Les cinnamyl alcool déhydrogénases (CAD)

La cinnamyl alcool déhydrogénase (CAD) est une enzyme multifonctionnelle et réalise la dernière étape de la voie de biosynthèse des monolignols en réduisant les aldéhydes en alcools. Contrairement au peuplier possédant une seule « vraie » enzyme CAD et 15 enzymes CAD-like (Tuskan *et al.*, 2006), la famille CAD chez *Arabidopsis* comprend sept gènes *CAD-like* et deux gènes codant pour des « vraies » CAD (Raes *et al.*, 2003 ; Eudes *et al.*, 2006). La famille CAD chez les plantes est divisée en quatre classes. La classe I comprend les gènes *AtCADD* et *AtCADC* d'*Arabidopsis* fortement homologues entre eux, et les gènes *CAD* de peuplier et de pin codant pour des « vraies » enzymes CAD impliquées dans la lignification. La classe II inclut les gènes *AtCADB.1* et *AtCADB.2* fortement similaires entre eux, le gène *AtCADA* et le gène *PtSAD* de peuplier. Les gènes *AtCAD1*, *AtCADE* et *AtCADF* sont groupés dans la classe III, et le gène *AtCADG* très divergent des autres membres de la famille CAD forme la classe IV. Les *CADD* et *CADC* d'*Arabidopsis* sont les enzymes de la famille CAD ayant la plus forte affinité pour les aldéhydes p-coumarylique, caffeylique, coniferylique, 5-hydroxyconiferylique et sinapylique. Les autres CAD n'ont qu'une faible affinité pour ces cinq substrats et pourraient réduire d'autres aldéhydes non liés à la synthèse de la lignine (Kim *et al.*, 2004). Ainsi, les protéines CAD groupées dans les classes II et III ne devraient pas être considérées comme des vraies CAD impliquées dans la lignification, mais plutôt comme des CAD-like ou des alcool déshydrogénases multifonctionnelles. En effet, beaucoup ont d'autres préférences de substrat, certaines sont des mannitol ou benzyle alcool déshydrogénases, et les mutants d'*Arabidopsis* correspondants n'ont pas un contenu ou une composition en lignine significativement altérée, excepté le mutant *Atcad1* (Raes *et al.*, 2003 ; Kim *et al.*, 2004 ; Eudes *et al.*, 2006). Concernant les gènes *AtCADC* et *AtCADD*, leur promoteur est actif dans les tissus vasculaires des feuilles et des

tiges où une forte activité est localisée dans les fibres interfasciculaires et les vaisseaux/fibres du xylème proches du cambium (Sibout *et al.*, 2003). Le gène *AtCADC* est également exprimé dans d'autres tissus comme les fleurs ou le parenchyme des feuilles, et seule l'expression du gène *AtCADD*, possédant un élément AC dans son promoteur, serait vraiment spécifique du xylème (Raes *et al.*, 2003 ; Sibout *et al.*, 2003). Bien que l'activité CAD soit fortement diminuée, surtout vis-à-vis des aldéhydes sinapyliques, dans chacun des mutants *cadd* et *cadc*, seul le mutant *cadd* montre une réduction légère de la teneur en lignine, et une proportion réduite d'unité S dans les tiges et racines s'accompagnant d'une forte incorporation d'aldéhyde sinapylique (Sibout *et al.*, 2003). De manière intéressante, le double mutant *caddcadc* présente un phénotype marqué avec une hampe florale avachie et une diminution de 40 % de la teneur en lignine dans les tiges (Sibout *et al.*, 2005). La teneur en unités guaiacyles et syringiques chute drastiquement (94 %) et il y a une forte incorporation d'aldéhydes synapylique et coniferylique dans la tige du double mutant. Par ailleurs, les gènes *CAD* de peuplier et de pin sont capables de restaurer le phénotype du double mutant (Sibout *et al.*, 2005). Enfin, le gène *AtCAD1* pourrait aussi être impliqué dans la lignification des tiges (Eudes *et al.*, 2006). En effet, une faible réduction du contenu en lignine est observée dans la tige en élévation du mutant *cad1* et le gène *AtCAD1* restaure partiellement le phénotype du double mutant *cadd cadc*. Toutefois, le gène *AtCAD1* ne possède pas d'élément AC dans son promoteur et est exprimé dans d'autres tissus que le xylème primaire comme par exemple les siliques (Raes *et al.*, 2003 ; Eudes *et al.*, 2006).

Chez le peuplier, les enzymes CAD et SAD sont respectivement groupés dans les classes I et II. La CAD serait plutôt spécifique de la transformation de l'aldéhyde coniferylique en alcool coniferylique pour la synthèse des unités G, tandis que la sinapyl alcool déshydrogénase (SAD) serait spécifique de la transformation de l'aldéhyde sinapylique en alcool sinapylique pour la synthèse des unités S (Li *et al.*, 2001). En effet, les aldéhydes coniferyliques sont des inhibiteurs compétitifs de la réduction des aldéhydes sinapyliques par la CAD, et inversement les aldéhydes sinapyliques inhibent fortement la réduction des aldéhydes coniferyliques par la SAD. De plus, chez le peuplier tremble, la protéine SAD est coexprimée avec la F5H et la COMT dans les cellules contenant de la lignine riche en unité S comme les fibres du phloème, et les cellules de rayon du xylème (Li *et al.*, 2001). Les protéines CAD et SAD sont respectivement plus abondantes aux stades précoce et tardif de la différenciation des fibres et des cellules de rayon du xylème. De plus, la CAD est fortement exprimée dans les éléments de vaisseaux contenant de la lignine enrichie en unité G. Toutefois, l'implication majeure de la SAD dans la synthèse d'unité S est remise en question

par la faible capacité de la SAD à restaurer des teneurs normales en unité S et en aldéhydes sinapyliques chez le double mutant *cadd cadc* d'*Arabidopsis* (Sibout *et al.*, 2005). Curieusement, chez des peupliers transgéniques montrant une forte réduction de l'activité CAD, il n'a été observé qu'une faible réduction de la teneur en lignine (Baucher *et al.*, 1996 ; Lapiere *et al.*, 1999). Par contre, ces peupliers présentent une modification de la structure des polymères de lignine, avec pour conséquence une meilleure extraction de cette lignine altérée par des agents alcalins. L'altération de l'activité de cette enzyme entraîne l'incorporation de précurseurs de monolignols tels que des aldéhydes hydroxycinnamiques dans les polymères de lignine ce qui modifie potentiellement le pouvoir de polymérisation des peroxydases et génère finalement un polymère de lignine plus fragmenté (Ralph *et al.*, 2001). Cette incorporation de cinnamaldéhydes expliquerait en partie la couleur rouge du xylème et la faible réduction de lignine chez ces peupliers transgéniques.

L'ensemble des résultats d'analyses de plantes transgéniques modifiées dans la biosynthèse des monolignols démontre que les plantes peuvent tolérer de larges variations dans la quantité et la composition en lignine, que des monomères inhabituels autres que les alcools *p*-coumarique, coniferylique et synapilique peuvent être incorporés dans la lignine résultant dans des structures nouvelles de lignine, et que finalement le polymère de lignine est très flexible dans sa composition (Boerjan *et al.*, 2003). Par ailleurs, il est important de signaler que certaines enzymes de la voie de biosynthèse des monolignols telles que par exemple les PAL, C4H, 4CL et C3H participent également à d'autres voies enzymatiques conduisant à la biosynthèse d'autres composés du métabolisme secondaire comme les flavonoïdes, anthocyanines, lignane, tanins et glycosides de phénols (Tsai *et al.*, 2006).

### 2.5.8. Régulation transcriptionnelle des gènes de la voie de biosynthèse des lignines

Le contrôle spatial temporel de la lignification est important pour le transport d'eau, la résistance aux maladies, le développement et le soutien de la plante. Beaucoup de gènes codant pour des enzymes impliquées dans la lignification sont régulés au niveau transcriptionnel (Rogers & Campbell, 2004). Nous avons vu précédemment que les gènes *PAL* répondent à différents stress et ont diverses spécificités d'expression tissulaire. Notamment le promoteur du gène *PAL* dirige l'expression du gène au niveau des cellules du xylème en différenciation suggérant une régulation transcriptionnelle du gène *PAL* dans le contrôle de la lignification (Bevan *et al.*, 1989 ; Ohl *et al.*, 1990). L'analyse du promoteur du

## Introduction Générale

gène *PAL2* de haricot a montré que certaines séquences régulatrices confèrent une spécificité d'expression tissulaire (Hatton *et al.*, 1995). Trois des motifs fonctionnant dans cette régulation sont des éléments AC riches en résidus adénine et cytosine, ne contenant pas de guanine. Les éléments AC-I et AC-II du promoteur *PAL2* de haricot sont nécessaires pour diriger l'expression du gène dans le xylème. Les éléments AC ont ainsi été subdivisés en boîtes ACI et ACII, bien que ces deux éléments s'alignent parfaitement et sont fonctionnellement redondants (Raes *et al.*, 2003). Les éléments AC ont aussi été appelés boîtes AC, boîtes H, ou encore boîte PAL (Douglas, 1996 ; Kawaoka & Ebinuma, 2001 ; Rogers & Campbell, 2004). Initialement découvert dans les promoteurs des gènes *PAL*, ces éléments AC ont aussi été identifiés dans les promoteurs d'autres gènes impliqués dans la biosynthèse de lignine, et ceci chez différentes plantes. Ce sont par exemple les gènes *C4H* d'*Arabidopsis*, *4CL* et *CCoAOMT* de persil, *CCR* et *CAD2* d'eucalyptus, et *prxC2* codant pour une peroxydase de raifort (Kawaoka & Ebinuma, 2001 ; Rogers & Campbell, 2004). Tous ces promoteurs contiennent un ou plusieurs éléments AC probablement suffisant pour diriger spécifiquement l'expression des gènes dans le xylème en différenciation et réprimer leur expression dans le phloème et les cellules du cortex. Toutefois, cette règle ne semble pas être stricte. En effet, le promoteur du gène *PAL2* de peuplier comportant deux éléments ACI et ACII putatifs est actif dans le xylème primaire des jeunes feuilles et tiges et dans les cellules épidermiques des feuilles, des tissus peu lignifiés (Gray-Mitsumune *et al.*, 1999). Chez *Arabidopsis*, sur les 34 gènes de la voie de biosynthèse des monolignols analysés, 14 gènes (*PAL1*, 2, 3 et 4, *C4H*, *4CL1* et 2, *HCT*, *C3H1*, *CCoAOMT1*, *CCR1*, *F5H1*, *COMT* et *CADD*) semblent potentiellement impliqués directement dans la biosynthèse de la lignine, de par leur classification phylogénétique et leur expression dans les tiges d'inflorescences (Raes *et al.*, 2003). De plus, parmi ces 14 gènes, l'implication de huit d'entre eux (*PAL1* et 2, *C4H*, *C3H1*, *CCR1*, *F5H1*, *COMT* et *CADD*) dans la biosynthèse de la lignine a été confirmée expérimentalement, et neuf gènes (*PAL1* et 2, *4CL1* et 2, *HCT*, *C3H1*, *CCoAOMT1*, *CCR1*, et *CADD*) contiennent au moins une copie de cet élément AC au sein des 1000 pb de la région 5' non codante (Raes *et al.*, 2003). Ainsi, cet élément AC commun a sûrement une importance fonctionnelle dans la plante, et serait un mécanisme permettant la régulation coordonnée dans les tissus vasculaires des gènes impliqués dans les différentes étapes de la biosynthèse de la lignine. De plus, tous les gènes de la voie de biosynthèse des monolignols contenant un élément AC dans leur promoteur codent pour des enzymes intervenant dans la branche des unités G de la voie de biosynthèse des lignines. La présence de cet élément AC est corrélé avec une forte expression des gènes impliqués dans la biosynthèse de lignine de type G, et

pourrait expliquer les différences spatiotemporelles dans la déposition des unités G et S dans les cellules du xylème secondaire des tiges (Raes *et al.*, 2003). Beaucoup de facteurs de transcription, notamment de type MYB et LIM, se liant à l'élément AC et régulant l'activité des promoteurs correspondants ont été identifiés et font encore l'objet de nombreuses études (Rogers & Campbell, 2004).

### 2.6. Les facteurs de transcription impliqués dans la formation de la paroi secondaire et la biosynthèse des lignines

Les facteurs de transcription sont des protéines se liant à l'ADN qui activent ou répriment l'expression des gènes dans une condition donnée. La régulation de la transcription des gènes est importante pour le développement de l'organisme et la réponse aux facteurs environnementaux. De plus, le génome d'*Arabidopsis thaliana* contient environ 1600 gènes codant pour des facteurs de transcription sur 25000 gènes (Riechmann *et al.*, 2000). Ces facteurs de transcription sont regroupés selon leur structure protéique en différentes familles multigéniques comme les familles bZIP, HD-Zip, MADS, EREBP, et bHLH. Certaines se sont spécialisées dans des fonctions métaboliques spécifiques de plante comme la famille MYB de type R2R3 ou la famille NAC (Martin & Paz-Ares, 1997 ; Demura & Fukuda, 2007).

#### 2.6.1. Les facteurs de transcription de type NAC et la formation de la paroi secondaire des fibres

Une nouvelle famille de facteurs de transcription très intéressante associée à la formation du xylème et à la croissance secondaire est la famille de facteurs de transcription à domaine NAC (pour NAM, ATAF1/2 et CUC2) (Demura & Fukuda, 2007). Les protéines NAC sont des facteurs de transcription spécifiques des plantes caractérisés par un domaine NAC N-terminal conservé, et une région divergente activatrice de la transcription en C-terminal (Ooka *et al.*, 2003). C'est une large famille comprenant respectivement 105 et 75 membres chez *Arabidopsis* et le riz. Elle est divisée en deux groupes majeurs et subdivisée en 18 sous groupes dont certains sont spécifiques de monocotylédones ou eudicotylédones. Sept gènes *VND1* à *7* pour « vascular-related NAC domain » ont une expression associée à la différenciation du xylème chez *Arabidopsis* (Kubo *et al.*, 2005). Parmi ces gènes, *VND6* et *VND7* montrent une expression spécifique des vaisseaux immatures du métaxylème et protoxylème. Ces deux gènes sont des facteurs de transcription régulant positivement la

## Introduction Générale

différenciation des vaisseaux du métaxylème et du protoxylème. Les gènes *NST1* et *NST2* pour « *NAC secondary wall thickening promoting factor* » sont essentiels pour la déhiscence normale des anthères chez *Arabidopsis* (Mitsuda *et al.*, 2005). La surexpression de ces deux gènes conduit à la formation ectopique de cellules ressemblant à des vaisseaux de la même manière que la surexpression des gènes *VND6* et *7*. Récemment, il a été montré que le gène *NST3* ou *SND1* pour « *secondary wall-associated NAC domain protein* » spécifiquement exprimé dans les fibres interfasciculaires et de xylème de la tige d'*Arabidopsis*, est un régulateur clé activant la synthèse de paroi secondaire des fibres (Zhong *et al.*, 2006). De manière importante, la surexpression du gène *SND1* induit une accumulation de cellulose, xylane et lignine, et une activation de l'expression de gènes impliqués dans la biosynthèse de la paroi secondaire. Il est notamment observé une très forte accumulation de transcrits codant pour les cellulose synthases secondaires *AtCesA7* et *AtCesA8*, les enzymes *CCoAOMT* et *4CL1*, et d'autres facteurs de transcription de type NAC, MYB et *KNAT7* (Zhong *et al.*, 2006). De plus la perte de fonction simultanée des gènes *SND1/NST3* et de *NST1* chez des mutants KO, des répresseurs dominants ou des lignées RNAi, résulte en la suppression totale de la formation de paroi secondaire dans les fibres, et la perte de cellulose, xylane et lignine (Mitsuda *et al.*, 2007 ; Zhong *et al.*, 2007). Ces deux gènes *NST1* et *NST3/SND1* sont donc redondants et semblent avoir une fonction pivot dans la régulation transcriptionnelle contrôlant la différenciation des fibres. Enfin, le gène *ANAC012* d'*Arabidopsis*, préférentiellement exprimé dans le procambium, pourrait agir comme un régulateur négatif de l'épaississement des parois secondaires des fibres du xylème (Ko *et al.*, 2007). La surexpression du gène *ANAC012* dans la plante supprime drastiquement la déposition de paroi secondaire dans les fibres du xylème et diminue la composition en cellulose des tiges et des racines. Ainsi, certains gènes NAC sont essentiels dans le contrôle de la formation de la paroi secondaire des fibres du xylème et les protéines NAC sont à ma connaissance les seuls facteurs de transcription susceptibles de réguler directement l'expression de gènes codant pour des cellulose synthases. Enfin, bien que potentiellement impliqué dans la formation des fibres de bois, aucun gène NAC n'a encore été étudié chez les arbres.



### 2.6.2. Les facteurs de transcription MYB impliqués dans la lignification

Beaucoup de facteurs de transcription de type R2R3-MYB sont capables de réguler la transcription de gènes impliqués dans la voie de biosynthèse des lignines. En effet, l'élément AC est reconnu par le domaine de liaison à l'ADN de certains facteurs de transcription R2R3-MYB. Contrairement aux autres eucaryotes, les facteurs de transcription MYB de plante sont pratiquement tous de type R2R3. Les protéines R2R3-MYB forment la plus grande famille de facteurs de transcription avec plus de 120 membres chez *Arabidopsis* (Riechmann *et al.*, 2000). Les protéines MYB contiennent un domaine N-terminal conservé liant l'ADN appelé domaine MYB, et un domaine C-terminal divergent qui fonctionne dans la régulation de la transcription (Martin & Paz-Ares, 1997). Les premiers gènes découverts codant pour des facteurs de transcription MYB régulant la lignification sont *AmMYB308* et *AmMYB330* du muflier (*Antirrhinum majus*). Ils sont capables de se lier sur un élément cis-activateur contenant l'élément AC (ou motif PAL-box) et régulent des gènes impliqués dans la voie de biosynthèse des phénylpropanoïdes et des lignines (Tamagnone *et al.*, 1998). La surexpression de l'un ou l'autre de ces gènes provoque une répression du métabolisme des acides phénoliques. De plus, les plantes surexprimant *AmMYB308* présentent une réduction d'au moins 17 % du contenu en lignine dans les tissus vasculaires et seule l'expression des gènes *C4H*, *4CL* et *CAD* est diminuée (Tamagnone *et al.*, 1998). Récemment, une étude a montré que le gène *PtMYB4* est exprimé spécifiquement dans les cellules lignifiées chez *Pinus taeda* (Patzlaff *et al.*, 2003). La protéine PtMYB4 est aussi capable de se fixer sur l'élément AC et d'activer la transcription à partir de l'élément AC chez la levure. Le gène *PtMYB4* est suffisant, quand il est surexprimé chez le tabac, à l'induction de la lignification dans les cellules non lignifiées et activerait l'expression des gènes *C3H*, *CCoAOMT*, *COMT*, *CAD*, et *CCR*. Cette induction ectopique de la lignification ne semble pas avoir d'effets sur l'accumulation d'autres classes de phénylpropanoïdes (Patzlaff *et al.*, 2003). De façon intéressante, il a été montré que cette protéine PtMYB4 est capable de fixer l'élément AC du promoteur du gène *PsPGS1b* codant pour une glutamine synthase chez le pin sylvestre (*Pinus sylvestris*) (Gomez-Maldonado *et al.*, 2004). Le gène *PsPGS1b* exprimé dans les tissus vasculaires jouerait un rôle dans le recyclage de l'ammonium produit par différents types de métabolismes azotés. Cette découverte montre que certains facteurs de transcription MYB relient et co-régulent le métabolisme azoté et la biosynthèse des lignines un processus nécessitant une forte demande en azote. Par ailleurs, *ALTERED PHLOEM DEVELOPMENT* (*APL*), un gène codant un facteur de transcription MYB chez *A. thaliana* et exprimé dans les

## Introduction Générale

tissus vasculaires, contrôle le développement vasculaire en induisant la différenciation du phloème, et inhibant celle du xylème (Bonke *et al.*, 2003). Ainsi, chez le mutant *apl* les cellules normalement positionnées au niveau du phloème présentent des caractéristiques de cellules du xylème, et la surexpression ectopique du gène *APL* dans les paquets vasculaires inhibe le développement du xylème. La protéine APL est un exemple de facteur de transcription MYB impliqué dans la différenciation des cellules vasculaires et ayant un effet indirect sur la lignification en inhibant la production de cellule de xylème. Chez le peuplier, trois gènes codant des protéines de type MYB ont été isolés et parmi ceux-ci, *PttMYB21a* est fortement exprimé dans les fibres de phloème et le xylème formant de la paroi secondaire (Karpinska *et al.*, 2004). Les peupliers portant une construction antisens du gène *PttMYB21a* ont une croissance réduite, un contenu en lignine soluble augmenté dans l'écorce, et une augmentation de l'expression du gène *CCoAOMT* dans le phloème. Ainsi, la protéine PttMYB21a serait peut être un répresseur de la transcription. La famille de facteur de transcription MYB chez les plantes est très grande. Les candidats MYB potentiellement impliqués dans la lignification des cellules de bois sont donc probablement très nombreux, d'où la difficulté de déterminer leurs rôles précis dans ce processus complexe nécessitant une régulation fine de l'expression des gènes. Néanmoins, beaucoup d'études ont mis en évidence l'importance de certains facteurs de transcription dans la régulation de processus physiologiques conduisant à la formation du bois, comme le développement vasculaire ou la biosynthèse des lignines. Par contre, à notre connaissance, aucun facteur de transcription régulant la formation du bois de tension n'a encore été découvert.

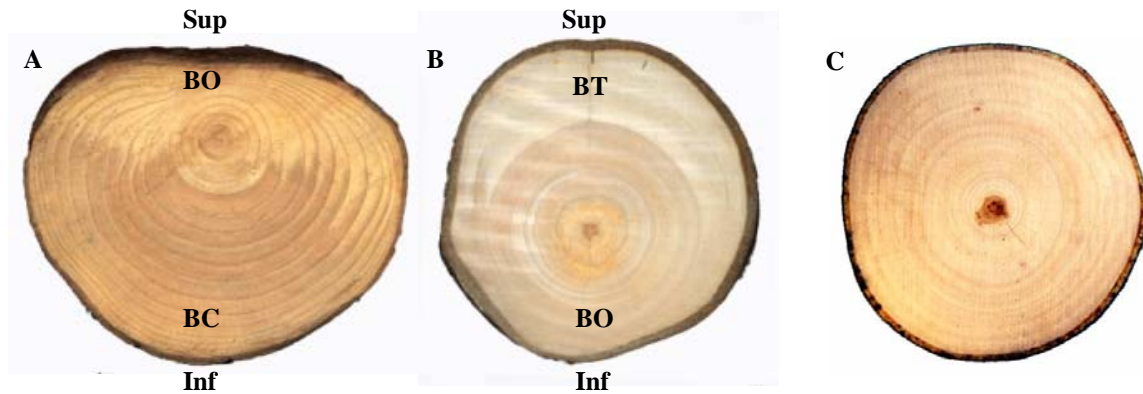
### 3 Le bois de tension, bois de réaction chez les feuillus

#### 3.1. Stimuli induisant la formation de bois de réaction

Le bois de réaction se forme en réponse à la perception de contraintes mécaniques et ou gravitationnelles comme le poids du houppier, le vent ou la pente (Scurfield, 1973). Ces contraintes induisent chez l'arbre une perturbation de la position d'équilibre de ces axes (tronc et branches). Le bois de réaction peut aussi être induit par des stimulus non gravitationnels, tels que la réorientation des branches latérales lorsque la branche principale (ou tronc) est éliminée, ou par le phototropisme. En effet, il a été montré chez le chêne *Quercus crispula* que le phototropisme induit la formation asymétrique de bois de tension pour une réorientation de la tige vers la source lumineuse (Matsuzaki *et al.*, 2007). En réponse à ces conditions défavorables de croissance, l'arbre s'adapte et forme du bois de réaction. Dans ce bois, d'importantes forces mécaniques sont mises en place et permettent ainsi la réorientation des tiges. Les tensions longitudinales de croissance produites dans le bois de réaction sont en partie le résultat d'une croissance radiale asymétrique (Fig. 19) due à une augmentation des divisions cellulaires des initiales cambiales couplée avec une différenciation particulière des cellules du xylème. Le bois de réaction est donc indispensable au bon développement de l'arbre, et permet à celui-ci de s'adapter et répondre efficacement aux contraintes physiques intrinsèques et environnementales.

On distingue deux types de bois de réaction : le bois de compression chez les conifères habituellement situé à la face inférieure des tiges inclinées, et le bois de tension chez les dicotylédones généralement situé à la face supérieure des tiges inclinées (Scurfield, 1973). Le bois situé à la face inférieure de la tige inclinée chez les dicotylédones et celui situé à la face supérieure de la tige inclinée chez les gymnospermes sont appelés bois opposé (Fig. 19). S'agissant du bois opposé des feuillus, certains auteurs font également référence à du bois de compression (Joshi, 2003). Bien que des forces de compression existent dans le bois opposé sous l'effet de la gravité, cette appellation peut prêter à confusion avec le bois de compression des conifères. Le bois des arbres droits (non inclinés) est considéré comme du bois « normal ». Toutefois, la présence de bois de tension est fréquemment observée dans du bois normal. En effet, les facteurs environnementaux ou intrinsèques comme le phototropisme, le vent ou le propre poids de l'arbre, exercent des stress mécaniques transitoires ou prolongés sur le tronc ou les branches qui doivent constamment réorienter leurs axes (Telewski, 2006).

De plus, la formation de bois de tension a déjà été observée sur des tiges verticales de peuplier ayant une croissance rapide (Isebrands & Benseid, 1972 ; Timell, 1986). C'est pourquoi, lors de l'étude des fibres de bois de tension, il est souvent préférable de les comparer à des fibres de bois opposé plutôt que des fibres de bois normal.



**Figure 19.** Coupe transversale de tronc montrant du bois de réaction chez un conifère et un feuillu (D'après la thèse de Lafarguette F., 2003).

(A) Coupe transversale de tronc de pin maritime (*Pinus pinaster*). Le bois de compression (BC) se forme sur la face inférieure (Inf) de la courbure chez les conifères.

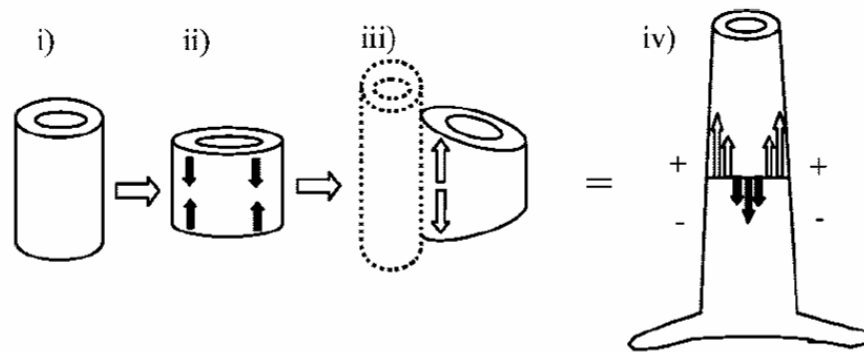
(B) Coupe transversale de tronc de peuplier incliné. Le bois de tension (BT) se forme sur la face supérieure (Sup) de la courbure chez les feuillus. Le bois de réaction se distingue du bois opposé (BO) s'apparentant à du bois normal à la face opposée. Les bois de compression et de tension sont caractérisés par une forte excentricité de la moelle respectivement aux faces inférieure et supérieure des troncs inclinés.

(C) Coupe transversale de tronc de peuplier droit

### 3.2. La différenciation du bois engendre des contraintes de maturation dans les tiges

Durant sa vie, l'arbre est soumis à un ensemble de forces internes qui va précontraindre son bois. Une partie de ces contraintes est due à des déformations de maturation qui apparaissent à la fin de la différenciation des cellules du xylème (Fournier *et al.*, 1994). Les cellules du xylème ont en effet tendance à se déformer durant leur maturation, et si elles étaient isolées, ces cellules subiraient une contraction longitudinale et une expansion transversale (Fig. 20). Or, dans le bois, les couches de cellules nouvellement formées sont déjà solidaires et collées aux couches de xylème plus internes et anciennes. Les déformations longitudinale et transversale des nouvelles cellules sont donc empêchées par la rigidité du bois, et ces cellules subissent alors une tension longitudinale et une compression transversale. Le bois plus interne va, lui, développer une compression longitudinale et une tension transversale pour équilibrer les deux types de forces (tension longitudinale et compression transversale) des cellules de la couche la plus externe. Ces contraintes résultant

de la croissance en diamètre du tronc sont appelées contraintes de maturation ou de croissance (Fournier *et al.*, 1994). De plus, la croissance intrusive des fibres au sein de la structure relativement rigide de la tige induit des forces compressives internes (Larson, 1994 ; Telewski, 2006). Le propre poids de l'arbre et la compression exercée par l'écorce sur le cambium, sont également des facteurs importants générant des contraintes mécaniques internes dans le tronc (Telewski, 2006). Les contraintes de croissance pourraient être générées par la mise en tension ou contraction des microfibrilles de cellulose cristalline (Bamber, 1987 ; Okuyama *et al.*, 1994).



**Figure 20.** Les contraintes de maturation dans le bois (Plomion *et al.*, 2001)

i) La nouvelle cellule de bois se développe, commence à se différencier et s'allonger. ii) Lors de la déposition de la cellulose et de la lignine dans la paroi secondaire, la cellule tend à s'étirer latéralement ce qui induit un rétrécissement longitudinal. iii) Or, la cellule de bois en différenciation est attachée aux couches de xylème plus âgées et ne peut se déformer complètement engendrant un stress mécanique dans la paroi cellulaire appelé contrainte de maturation. iv) Dans le bois normal, cette contrainte se traduit par un stress de tension, de telle manière à ce que le bois de la surface externe est habituellement tenu par des tensions (+). Toutefois, chaque nouvelle couche de cellules formée exerce une force tangentielle sur le bois déjà présent dans le tronc. Le bois interne subit donc une compression (-).

Les contraintes de maturation peuvent varier sur la circonférence de l'arbre notamment lors de la formation de bois de réaction. Les forces mécaniques engendrées localement sont différentes selon que l'espèce est un conifère ou un feuillu. Dans les zones de bois de réaction, il existe d'une part un pic de contraction longitudinale dans le bois de tension des feuillus où la contrainte initiale est une tension longitudinale, et d'autre part un pic de dilatation longitudinale dans le bois de compression des conifères où la contrainte initiale est une compression longitudinale (Fournier *et al.*, 1994). La formation de bois de tension est donc associée à de fortes valeurs en déformation de maturation. La présence de bois de tension dans l'arbre va induire une forte dissymétrie dans la répartition des contraintes de croissance permettant ainsi de modifier mécaniquement l'orientation des différents axes dans l'espace en exerçant une force de traction. Par ailleurs, la présence de forte quantité de bois de tension dans l'arbre déprécie la qualité du bois. En effet, celui-ci entraîne des défauts

technologiques comme l'apparition de fentes à l'abattage, l'aspect pelucheux du bois au niveau des faces de sciage, le retrait au séchage et l'apparition de fentes au déroulage.

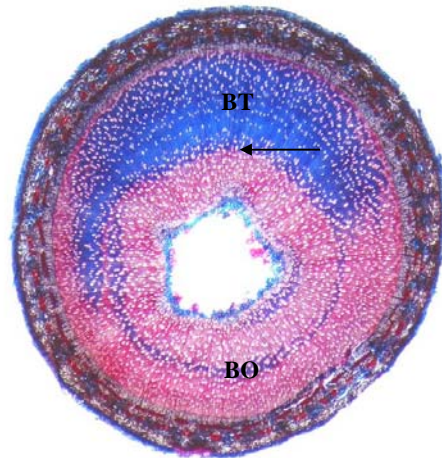
### 3.3. *Le bois de compression chez les conifères*

Le bois de réaction chez les résineux et les feuillus se caractérise principalement par des parois cellulaires ayant des propriétés structurales, biochimiques et mécaniques particulières. Chez les conifères, le bois de compression situé à la face inférieure des tiges inclinées est caractérisé par une excroissance radiale résultant d'une forte activité cambiale (Fig. 19). Le bois de compression est aussi caractérisé par une couleur rouge due soit à la polymérisation de coniféraldéhyde dans le polymère de lignine, soit à la biosynthèse de flavonoïdes et l'accumulation de pigments anthocyanines (Higuchi *et al.*, 1994 ; Gion *et al.*, 2005). Mécaniquement, ce bois est caractérisé par une contrainte longitudinale de compression qui s'oppose à la contrainte de tension du bois normal formé à la face supérieure. Cette force de compression réoriente les tiges en les « poussant » vers le haut (Plomion *et al.*, 2001). Dans ce bois particulier, la morphologie des cellules est modifiée : les trachéides sont plus courtes, arrondies et possèdent une paroi cellulaire plus épaisse. Cet épaissement des parois est responsable de la densité élevée du bois de compression, supérieure de 50 % par rapport à celle de bois normal. La forme arrondie des trachéides laisse apparaître des espaces intercellulaires qui sont plus larges que dans le bois normal. De plus, les rayons semblent élargis transversalement (Hellgren *et al.*, 2004). A l'inverse du bois de tension, la teneur en cellulose diminue dans le bois de compression. L'angle des microfibrilles de cellulose dans la couche S2 est plus élevé dans le bois de compression (30° à 45° par rapport à l'axe longitudinal des trachéides) que dans le bois normal (5° à 30°). De plus, le bois de compression est fortement lignifié au niveau de sa couche S2 (aussi appelée couche S2L) et il contient plus d'unités *p*-hydroxyphényle (H) que le bois normal (Donaldson, 2001). Ce fort dépôt de lignine entre les microfibrilles de cellulose serait responsable de l'augmentation de l'angle des microfibrilles dans la couche S2.

### 3.4. Caractéristiques morphologiques du bois de tension

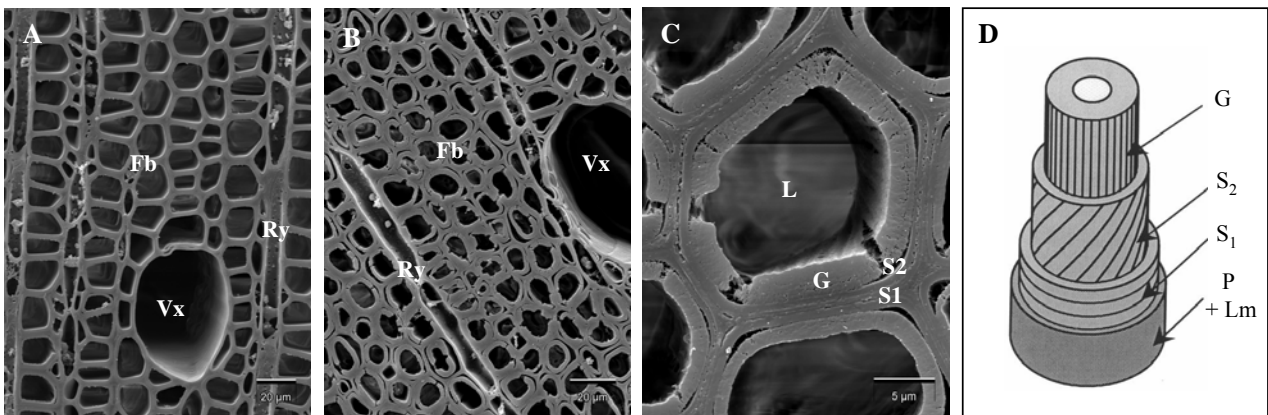
Le bois de tension se distingue du bois normal par ses caractéristiques anatomiques, chimiques, physiques et mécaniques. L'anatomie du bois de tension varie beaucoup en fonction des espèces de feuillus. La première différence est l'excentricité de la moelle due à une importante activité cambiale au niveau de la partie en tension (Fig. 19 et 21). Cette excentricité de la moelle peut être due à une augmentation des divisions cellulaires du cambium, une expansion radiale accrue des fibres et des vaisseaux, ou une durée de différenciation plus longue des cellules cambiales vers les fibres matures. Comme les fibres de bois normal, les fibres gélatineuses sont plus larges et longues que les cellules initiales fusiformes et portent de simples ponctuations (Chaffey *et al.*, 2002). Dans cette étude histologique des différentes cellules de bois chez le peuplier, il n'a pas été mentionné de différences dans la longueur des fibres entre bois tendu et bois normal. Toutefois, les résultats d'une étude effectuée chez l'hybride *Populus euramericana* montrent que les vaisseaux et les fibres de bois tendu ont un diamètre légèrement rétréci, et sont respectivement plus longs de 10 et 27  $\mu\text{m}$  par rapport à ceux de bois opposé (Jourez *et al.*, 2001). Il est à préciser que cette différence dans la longueur de ces cellules est plutôt faible comparée à la longueur moyenne des fibres (610  $\mu\text{m}$ ) et des vaisseaux (385  $\mu\text{m}$ ). Cette différence de longueur peut-être due à une prolongation de la différenciation des fibres dans le bois de tension. De façon plus marquée, le nombre de vaisseaux dans le bois de tension diminue nettement avec une fréquence en vaisseaux 33 % plus faible que dans le bois opposé. En contrepartie, la proportion des fibres augmente dans le bois de tension. Il en résulte une diminution de 29 % de la porosité dans le bois de tension par rapport au bois opposé. Bien que cela n'ait pas été directement démontré, l'excentricité de la moelle pourrait plutôt résulter d'une forte activité mitotique des initiales fusiformes (Jourez *et al.*, 2001 ; Hellgren *et al.*, 2004). Contrairement au bois de compression où les trachéides prennent un aspect arrondi, la forme des différentes cellules (fibres, vaisseaux et rayons) ne semble pas être modifiée dans le bois de tension. Enfin, en coupe transversale, les cellules des rayons de bois de tension apparaissent globalement plus nombreuses et plus courtes que dans le bois opposé (Jourez *et al.*, 2001).

## Introduction Générale



**Figure 21.** Le bois de tension contient plus de cellulose et moins de lignines comparé au bois opposé (Photo Laurans F., INRA Orléans).

Coupe transversale d'une jeune tige inclinée de peuplier (*P. tremula* × *P. alba*) colorée à la safranine bleu-astre. Les colorations bleu et rouge révèlent respectivement la cellulose et la lignine. Comparé au bois opposé à la face inférieure (**BO**), le bois tendu à la face supérieure (**BT**) fortement coloré en bleu est enrichi en cellulose. Notons l'excentricité de la moelle due à une forte activité cambiale dans le bois de tension, et le moment de l'inclinaison de la tige induisant la formation de bois de tension (marqué par une flèche).



**Figure 22.** Les fibres de bois tendu chez le peuplier (*P. tremula* × *P. alba*) (Photos Laurans F., INRA Orléans).

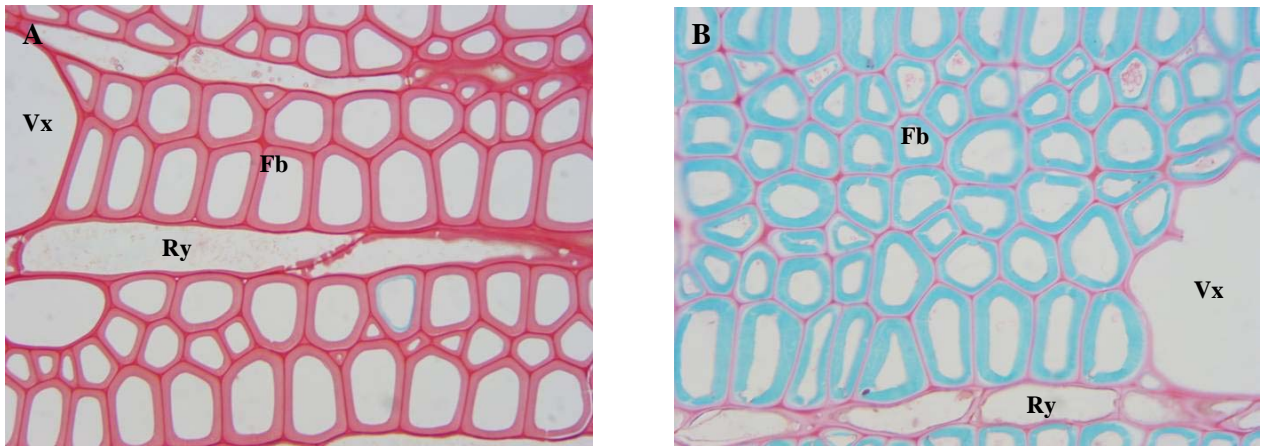
(**A** et **B**) Photos en microscopie à balayage de coupes transversales de bois opposé (**A**) et de bois tendu (**B**). Trois types cellulaires sont visibles : les fibres (**Fb**), les vaisseaux (**Vx**) et les cellules de rayon (**Ry**). Seules les fibres possèdent une couche supplémentaire d'aspect gélatineux. (**C**) Coupe transversale de fibres de xylème de bois tendu en microscopie électronique à balayage. Les fibres présentent une couche gélatineuse caractéristique appelée couche **G**, ainsi qu'une couche **S2** moins épaisse, et une couche **S1** très fine. **L** représente la lumière de la fibre, sans contenu cellulaire. (**D**) Représentation schématique d'une fibre de xylème de bois tendu avec les différentes couches de la paroi et l'orientation des fibres de cellulose. De l'extérieur vers l'intérieur, on a la lamelle moyenne (**Lm**), la paroi primaire (**P**), et la paroi secondaire composée de la succession des couches **S1**, **S2** et une couche **S3** remplacée par la couche **G** chez le peuplier. L'angle des microfibrilles de cellulose est très faible dans la couche **G** et celles-ci sont presque alignées parallèlement à l'axe de la fibre.



### 3.5. *La paroi des fibres de bois tendu est caractérisée par la présence d'une couche gélatineuse chez le peuplier*

Les parois cellulaires des fibres de bois de tension ont des propriétés structurales, biochimiques et mécaniques particulières. Chez le peuplier et beaucoup d'autres arbres feuillus, la différence majeure du bois de tension par rapport au bois opposé est la présence, à la place ou en plus de la couche S3, d'une couche supplémentaire d'aspect gélatineux et translucide au niveau de la face interne de la paroi des fibres de xylème (Fig. 22). Chez le peuplier (*P. deltoides* Bartr. ex Marshall), le développement de cette couche gélatineuse semble affecter le développement de la couche S2 laquelle est plus fine dans la fibre gélatineuse que dans celle de bois « normal » (Joseleau *et al.*, 2004). Certaines espèces de feuillus, comme les *Magnolia* des angiospermes primitifs, ne possèdent pas de couche gélatineuse supplémentaire, mais peuvent développer au niveau du bois de tension une couche S2 ayant des caractéristiques apparentées à cette couche pariétale particulière (Yoshizawa *et al.*, 2000). Cette couche gélatineuse, appelée couche G, n'est pas visible au niveau des vaisseaux et des cellules des rayons, et a également été observée dans les fibres de phloème (Timell, 1986). Le bois opposé est totalement dépourvu de fibres gélatineuses (Fig. 21 et 23). Cette couche G est très épaisse chez le peuplier et est peu adhérente aux autres couches de la paroi (Fig. 22). Elle est constituée en grande majorité (95 %) de microfibrilles de cellulose très cristalline, et cette caractéristique est responsable de l'augmentation importante de la teneur en cellulose dans le bois de tension par rapport au bois opposé. L'angle des microfibrilles de cellulose dans la couche G est très faible et celles-ci sont organisées presque parallèlement à l'axe de la fibre (Fig. 22). Sur une coupe transversale de tige de peuplier inclinée et colorée à la safranine bleu astra, la forte teneur en cellulose, apparaissant en bleu, est très marquée dans le bois de tension (Fig. 21 et 23). Cette couche G semble être responsable des propriétés mécaniques particulières de ce bois. Une étude originale effectuée sur différentes espèces de feuillus formant ou non des fibres G dans le bois de tension a montré que d'importantes contraintes de tension longitudinale apparaissent au niveau des zones où les couches G sont abondantes (Okuyama *et al.*, 1994). De plus, pour les espèces ne formant pas de couche G, les corrélations suivantes ont été observées : Les contraintes de tension longitudinale sont d'autant plus grandes que la teneur en cellulose et sa cristallinité augmentent, et que l'angle des microfibrilles de cellulose est petit. Ces résultats montrent l'importance des microfibrilles de cellulose dans l'établissement des contraintes de maturation dans le bois de tension. Enfin, des expériences récentes de diffraction aux rayons X sur du

bois de tension de peuplier, supportent fortement l'hypothèse que la tension est induite dans les microfibrilles de cellulose (Clair *et al.*, 2006).



**Figure 23.** Les bois de tension et opposé chez le peuplier (*P. tremula* × *P. alba*) en microscopie optique (Photos Laurans F., INRA Orléans).

Coupes transversales radiales de bois opposé (A) et de bois tendu (B) colorées à la safranine bleu astra. Trois types cellulaires sont visibles : les fibres (Fb), les vaisseaux (Vx) et les cellules de rayons (Ry). Les parois des fibres de xylème tendu possèdent une couche gélatineuse interne enrichie en cellulose et fortement colorée en bleu.

### 3.6. Le bois de tension contient moins de lignine

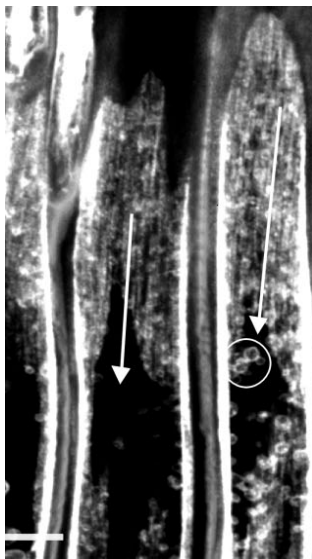
A la fois pour les conifères et les dicotylédones, il y a une plus forte lignification dans le bois à la face inférieure par rapport à la face supérieure de la tige inclinée. La lignine est donc un élément important lors de la formation de bois de réaction. Le bois de tension contient généralement moins de lignine et plus de cellulose, avec une augmentation générale de la teneur en cellulose dans le bois de tension de 10 à 20 % en fonction de l'espèce et une diminution égale en lignine et hémicellulose (Timell, 1969 ; Wada *et al.*, 1995). Mais, alors que les éléments de vaisseaux sont bien lignifiés, la diminution de la teneur en lignine dans le bois tendu est exclusivement portée par les fibres formant une couche G enrichie en cellulose (Fig. 23). En effet, les couches S1 et S2 des fibres de bois tendu contiennent au moins autant de lignine et de xylanes que celles de bois normal et de plus, ces deux couches sont plus fines que celles de bois normal. Cette diminution de la lignification dans le bois de tension est donc une caractéristique exclusive de la couche G (Timell, 1969). Les colorations avec le double colorant safranine bleu astra montrent que cette couche G est très peu lignifiée, alors que les autres couches S1 et S2 apparaissent fortement lignifiées (Fig. 23). Chez l'eucalyptus (*Eucalyptus viminalis*), la teneur en lignine ne varie pas dans le bois opposé avec la durée

d'inclinaison, alors que dans le bois tendu la teneur en lignine diminue après 10 jours d'inclinaison (Aoyama *et al.*, 2001). Toutefois, une faible teneur en lignine a bien été constatée dans cette couche gélatineuse par des observations immunohistochimiques et des mesures spectroscopiques (Yoshida *et al.*, 2002 ; Joseleau *et al.*, 2004). La composition de la lignine en sous unité guaiacique G et syringique S est particulière dans le bois de tension et il a été reporté une augmentation du ratio S/G dans le bois de tension. Chez l'eucalyptus, le ratio S/G est plus élevé dans le bois de tension que dans le bois opposé ou normal (Aoyama *et al.*, 2001). De plus l'activité des peroxydases pariétales catalysant préférentiellement l'oxydation des unités syringiques augmente dans le bois de tension. Chez le peuplier, des analyses immunohistochimiques utilisant des anticorps dirigés contre des épitopes de lignine GS, G et S montrent que les sous unités G et S de lignines sont présentes dans les fibres gélatineuses de bois tendu, mais en faible quantité. De plus ces analyses confirment que ces quelques polymères de lignines sont enrichis en sous unités S (Joseleau *et al.*, 2004). De façon intéressante, dans les fibres normales et gélatineuses le marquage des unités S est plus fort à la face interne de la couche G ou S2 qu'à la face externe indiquant que les unités syringiques de la lignine sont déposées en plus grande quantité lors des phases tardives de la formation de la paroi secondaire. Des observations similaires ont été réalisées par microscopie Raman confocale sur du bois de tension où un spectre correspondant à de la lignine est détecté au niveau d'une fine bande (0,5  $\mu\text{m}$ ) à la face interne de la couche G (Gierlinger & Schwanninger, 2006). Par contre, la face de la couche G accolée à la couche S2 semble totalement dépourvue de lignine.

### *3.7. Organisation du cytosquelette dans les cellules de bois de tension*

Il est très intéressant de constater que le cytosquelette dans les fibres de bois tendu présente un arrangement différent comparé au bois normal (Chaffey *et al.*, 2002). Notamment, les microtubules corticaux changent radicalement d'orientation dans les fibres gélatineuses. Toujours bien ordonnés parallèlement entre eux, ils deviennent arrangés de manière axiale dans le sens longitudinal de la fibre G (Fig. 24). Les microfilaments d'actine gardent leurs orientations axiales comme dans les fibres de bois normal, et se retrouvent co-alignés avec les microtubules. Au final, ces deux composants du cytosquelette se retrouvent donc approximativement alignés parallèlement aux microfibrilles de cellulose de la couche G. De par le changement d'orientation des microtubules similaire à celui des microfibrilles de

cellulose de la couche G, ces résultats suggèrent fortement que les microtubules sont impliqués dans la biosynthèse des microfibrilles de cellulose et contrôlent leur orientation dans le bois tendu (Chaffey *et al.*, 2002). L'organisation des microtubules et microfilaments d'actine au niveau des différentes ponctuations ne varie pas entre le bois normal et le bois tendu, et ne semble donc pas être affectée par l'inclinaison des tiges. Par ailleurs, aucune modification de l'organisation du cytosquelette n'a été reportée pour les autres cellules du bois de tension (éléments de vaisseaux, cellules axiales parenchymateuses et cellules de rayons).



**Figure 24.** Organisation des microtubules dans les fibres de bois de tension chez le peuplier *P. tremula* × *P. tremuloides* (Chaffey *et al.*, 2002) Les fibres gélatineuses vues en coupe longitudinale radiale. Les microtubules sont détectés par immunofluorescence indirecte de l' $\alpha$ -tubuline et visualisés en microscopie confocale laser. Les microtubules denses et indiqués par des flèches, sont alignés parallèlement et axialement au sein de la cellule. La région encerclée montre la présence de grains d'amidon dans la fibre G. (Échelle : 10  $\mu$ m)

### 3.8. Le bois de tension, facilement inductible est un bon modèle d'étude

La formation de bois de tension possédant des caractéristiques différentes au bois opposé ou « normal » est facilement inductible par la simple inclinaison des tiges de peuplier en croissance, et ce de manière reproductible (Pilate *et al.*, 2004b). Habituellement, il existe deux méthodes d'inclinaison des jeunes pousses pour l'induction expérimentale de la formation de bois tendu et opposé. Soit les plants inclinés se redressent librement sans contraintes pour revenir à leurs positions d'origine, soit les plants inclinés sont maintenus par des tuteurs pour forcer la formation de bois de tension sur toute la longueur des tiges et maintenir la présence de ce bois tout au long de l'expérience (Fig. 25). C'est cette dernière méthode qui est généralement appliquée au laboratoire. Sachant que la quantité de fibres G augmente en fonction de l'intensité du stimulus (Jourez *et al.*, 2003), les arbres sont fortement inclinés à 45° par rapport à la verticale et sont maintenus ainsi pendant environ un mois,

période à laquelle la quantité de fibre G est conséquente (Jourez & Avella-Shaw, 2003), et suffisante pour des analyses moléculaire et biochimique. Par ailleurs, la formation de bois de tension caractérisée par l'apparition d'une excentricité de la moelle et de fibres gélatineuses, est facilement contrôlable par des observations anatomiques et des colorations histochimiques. Notamment, la safranine bleu astra colorant la cellulose en bleu et les lignines en rouge, est couramment utilisée au laboratoire pour vérifier la présence de la couche G au sein des fibres de bois tendu. Ainsi, le bois de tension peut nous servir de modèle pour l'étude de la croissance radiale du cambium et de la formation des fibres gélatineuses. C'est en effet un bon système expérimental pour analyser les processus de différenciation des cellules du xylème et comprendre les voies biochimiques contrôlant le flux de carbone vers la cellulose, l'hémicellulose et la lignine.



**Figure 25.** Dispositif expérimental d'inclinaison des peupliers pour induire la formation de bois de tension

Les peupliers, clone INRA 717-1-B4 (*Populus tremula* × *P. alba*) sont inclinés à 45° par rapport à la verticale par inclinaison des pots, et les tiges sont maintenues par un tuteur.

### 3.9. Modification de l'expression des gènes dans le bois de tension

Ces modifications anatomiques et biochimiques au sein du bois de tension résulteraient en partie de la régulation coordonnée de l'expression d'un ensemble de gènes. En effet, le séquençage d'ESTs à partir de bois de tension et opposé et l'analyse de leurs distributions entre les différentes banques d'ADNc révèlent de nombreuses différences dans l'abondance de certaines ESTs dans ces deux types de bois (Déjardin *et al.*, 2004). Si l'on considère l'ensemble des gènes analysés, l'expression globale des gènes augmente dans le bois tendu, et diminue dans le bois opposé. Autrement dit, peu de gènes ont une expression induite dans le bois opposé, alors que l'expression d'un nombre significatif de gènes est

induite dans le bois de tension. Cette observation rend compte de la forte activité cambiale et de l'augmentation de la croissance radiale du bois de tension. Plus tard, les études transcriptomiques réalisées chez le peuplier et l'eucalyptus ont donné plus de précisions sur l'induction ou la répression de certains gènes dans le bois de tension (Paux *et al.*, 2005 ; Andersson-Gunneras *et al.*, 2006).

### 3.9.1. Les protéines fascicline-like à arabinogalactanes (FLAs) seraient des protéines pariétales essentielles à la formation de bois de tension

Certaines séquences codant pour des arabinogalactanes protéines (AGPs) sont spécifiquement induites dans le bois de réaction (Zhang *et al.*, 2000 ; Déjardin *et al.*, 2004 ; Lafarguette *et al.*, 2004 ; Andersson-Gunneras *et al.*, 2006). Chez le pin (*Pinus taeda*), plusieurs gènes codant pour des protéines à arabinogalactanes sont fortement exprimés dans les tissus vasculaires en différenciation formant de la paroi secondaire (Loopstra & Sederoff, 1995 ; Zhang *et al.*, 2000 ; Zhang *et al.*, 2003). Les analyses de distribution d'ESTs montrent qu'au moins trois AGPs sont surreprésentées dans le bois de compression (Whetten *et al.*, 2001). Les AGPs sont des glycoprotéines fortement glycosylées et excrétées au niveau du plasmalemme. Elles semblent être impliquées dans de nombreuses fonctions comme, le développement cellulaire, la reconnaissance cellulaire et la mort cellulaire programmée (Showalter, 2001). Chez le peuplier, quinze gènes codant pour des AGPs ont été appelées fasciclin-like AGP (PopFLA) et contiennent deux domaines AGP-like, un domaine fasciclin-like, et pour neuf d'entre eux, une ancre glycosylphosphatidylinositol (GPI) (Lafarguette *et al.*, 2004). Dix gènes *PopFLA1* à *10* sont fortement induits dans le bois de tension et ne sont pas exprimés dans la zone cambiale indiquant leur rôle dans les processus tardifs de la formation de la paroi secondaire. Cette forte induction des gènes codant pour des *PopFLAs* dans le bois de tension a été confirmée par une récente analyse globale de l'expression des gènes dans le bois de tension (Andersson-Gunneras *et al.*, 2006). Les gènes *PopFLA11* à *15* ne sont pas induits dans le bois de tension, et codent pour des protéines légèrement différentes dans leurs structures avec un domaine AGP-like N-terminal plus court que celui des protéines *PopFLA1* à *10* (Lafarguette *et al.*, 2004). Ces différences structurales pourraient expliquer des différences fonctionnelles dans la formation du bois de tension. De plus, les analyses immunohistochimiques montrent que les AGPs sont bien des protéines pariétales et sont localisées à la face interne de la couche G des fibres de xylèmes. Des résultats préliminaires d'immunohistolocalisation confirment la localisation de certaines *PopFLA* au niveau des

fibres G du bois tendu (Takeuchi M., communication personnelle). Ces éléments suggèrent l'importance de ces protéines dans l'élaboration de la couche G (Lafarguette *et al.*, 2004). Par ailleurs, la forte expression de certaines AGPs dans le bois de compression et dans la partie supérieure de la tige inclinée chez *Arabidopsis* (Whetten *et al.*, 2001 ; Yokoyama & Nishitani, 2006), des tissus structurellement et biochimiquement très différents du bois de tension, suggère que ces protéines sont plus généralement impliquées dans la réponse du bois de réaction au stress mécanique ou de la tige au gravitropisme. Cette dernière hypothèse a déjà été soulevée auparavant (Baluska *et al.*, 2003), et il a en effet été suggéré que ces AGPs pariétales soient des mécanorécepteurs interagissant avec le cytosquelette. Une étude récente montre en effet que dans les cellules « *Bright yellow 2* » (BY2) de tabac, les AGPs pourraient organiser les microtubules corticaux et induire des microfilaments d'actines plus épais (Sardar *et al.*, 2006). De la même manière, les microtubules et les microfilaments d'actines auraient un rôle dans l'organisation des AGPs à la surface cellulaire et au niveau des « *Hechtian strands* » qui relie la membrane à la paroi chez la plante.

### 3.9.2. Induction dans le bois de tension de l'expression des gènes impliqués dans la biosynthèse de cellulose

L'importante quantité de cellulose dans les fibres de bois de tension suppose une forte régulation de l'expression des gènes impliqués dans le métabolisme glucidique et dans l'élaboration des microfibrilles de cellulose. En effet, plusieurs gènes impliqués dans la biosynthèse des hydrates de carbones et de la cellulose comme certaines sucrose synthase, fructokinase et cellulose synthases sont induits dans le bois de tension (Déjardin *et al.*, 2004 ; Andersson-Gunneras *et al.*, 2006). Les cellulose synthases utilisent comme substrat l'UDP-glucose pour élaborer les polymères de glucose. L'UDP-glucose (et le fructose) sont produits par les sucrose synthases (SUS) à partir d'UDP, et de saccharose. Il a en effet été observé dans le xylème de bois tendu, une très forte induction de l'expression des gènes *PtSUS1* et 2 codant pour des sucrose synthases et une diminution concomitante de saccharose comparé au bois normal (Andersson-Gunneras *et al.*, 2006). Ces résultats indiquent que le saccharose est utilisé par les protéines SUS1 et 2 afin de fournir suffisamment de glucose aux cellulose synthases pour la synthèse de cellulose.

Les trois gènes *PtrCesA1*, *PtrCesA2* et *PtrCesA3* codant pour des cellulose synthases secondaires chez le peuplier (*P. tremuloides*) sont induits dans le bois de tension par rapport au bois opposé (Joshi, 2003). Comme le montrent des récentes expériences d'hybridation *in*

## Introduction Générale

*situ*, l'expression de ces trois gènes est fortement réprimée dans le bois opposé et est induite à la face supérieure (tendu) des tiges de peuplier inclinées. Cette induction est restreinte au xylème en développement et aux fibres de phloème (Bhandari *et al.*, 2006). L'expression du gène *PttCesA3-2* (l'homologue de *PtrCesA1*) est fortement induite dans le xylème de bois de tension par rapport à du bois normal, et l'expression du gène dupliqué *PttCesA3-1* l'est également mais de façon bien moins prononcée (Djerbi *et al.*, 2004 ; Andersson-Gunneras *et al.*, 2006). Par ailleurs, l'utilisation du promoteur de la cellulose synthase de peuplier (*PtrCesA1*) fusionné au gène rapporteur GUS en expression hétérologue chez le tabac, montre que l'expression du gène *PtrCesA1* est induite uniquement dans les fibres de phloème sous l'effet d'un stress de tension et est réprimée dans les tissus du côté opposé à la force de tension (Wu *et al.*, 2000). L'expression du gène *PttCesA8-2* serait elle aussi légèrement induite dans le bois de tension (Andersson-Gunneras *et al.*, 2006). A l'inverse et de façon surprenante en regard de la composition biochimique du bois opposé, il existe une cellulose synthase (*PttCesA2*) dont l'expression du gène serait trois fois plus élevée dans le bois opposé par rapport au bois tendu, uniquement 21 jours après inclinaison des tiges (Djerbi *et al.*, 2004). Lors d'une cinétique de formation de bois de tension pendant une semaine chez l'eucalyptus, l'expression d'une cellulose synthase probablement impliquée dans la formation de la paroi secondaire, est graduellement induite dans le bois de tension (Paux *et al.*, 2004 ; Paux *et al.*, 2005). Cette induction de l'expression d'un gène *CesA* secondaire dans le bois de tension d'eucalyptus, une espèce ligneuse phylogénétiquement éloignée du peuplier, confirme l'importance des cellulose synthases secondaires lors de la formation des fibres G de bois de tension.

Par ailleurs, les gènes *CesA* secondaires ne sont pas les seuls gènes associés à la biosynthèse de cellulose ayant une expression induite dans le bois de tension. En effet, il a été récemment montré par des analyses d'hybridation *in situ* que le transcrit *PtrKOR* et la protéine correspondante sont largement plus abondants dans le xylème en développement et le phloème tendu que dans le bois opposé (Bhandari *et al.*, 2006). Le gène *KORRIGAN* et les gènes *PtrCesA1* à 3 de peuplier ont ainsi un profil d'expression similaire caractérisé par une induction dans le bois de tension et une répression de l'expression dans le bois opposé. Ces résultats suggèrent fortement que ces quatre protéines coopèrent ensemble dans l'augmentation de la teneur en cellulose de la couche G du bois de tension (Bhandari *et al.*, 2006). Par ailleurs, un gène codant pour un homologue aux protéines COBRA est fortement induit dans le bois de tension (Andersson-Gunneras *et al.*, 2006). Chez *Arabidopsis*, COBRA est une protéine associée par une ancre glycosylphosphatidylinositol (GPI) à la surface



extracellulaire au niveau des parois longitudinales des cellules de racine, contrôlant l'expansion anisotrope des cellules dans beaucoup d'organes et jouant un rôle dans la déposition de la cellulose (Schindelman *et al.*, 2001 ; Roudier *et al.*, 2005). En effet, les mutants *cob*, *cob-1* et *cob-4*, touchant le gène *COBRA*, présentent une mauvaise orientation de l'expansion cellulaire, une désorganisation des microfibrilles de cellulose et en conséquence une diminution de la quantité de cellulose cristalline dans la zone de croissance racinaire. De façon intéressante, la localisation de la protéine COBRA pendant l'élongation cellulaire est dépendante de l'organisation des microtubules corticaux (Roudier *et al.*, 2005).

### 3.9.3. Expression des gènes impliqués dans l'organisation du cytosquelette dans le bois de tension

En relation avec le changement d'orientation des microfibrilles de cellulose dans la couche G de bois tendu, le cytosquelette et en particulier le réseau de microtubules corticaux a probablement une importance cruciale dans le contrôle de l'orientation des microfibrilles. En effet, les microtubules sont orientés longitudinalement et presque co-alignés avec les microfilaments d'actine et les microfibrilles de cellulose (Chaffey *et al.*, 2002), et certains gènes codant pour des isoformes de tubuline comme *TUB6* sont fortement exprimés dans le bois de tension (Andersson-Gunneras *et al.*, 2006). De plus, la quantité de transcrite d'un gène apparenté à une protéine motrice kinésine, liée aux microtubules, est augmentée dans le bois de tension chez l'eucalyptus (Paux *et al.*, 2005). Chez le peuplier, sur les six transcrits homologues aux kinésines déposés sur la lame de microarray, deux ont une expression induite dans le bois de tension, et les quatre autres sont réprimés (Andersson-Gunneras *et al.*, 2006). Ces protéines apparentées aux kinésines et induites dans le bois de tension pourraient avoir un rôle dans l'orientation des microfibrilles de cellulose, comme il a été suggéré pour le gène *FRA1* codant pour une kinésine et dont le mutant *fra1* présente une diminution de la solidité des tiges et une altération dans la déposition des microfibrilles de cellulose (Zhong *et al.*, 2002). Par contre, l'expression d'un gène codant pour la chaîne légère de la dynéine, une autre protéine motrice guidée par les microtubules, est légèrement réprimée dans le bois de tension. Enfin, il est remarquable de constater que l'expression de l'homologue du gène *AtKTNI* codant pour une katanine est également induite dans le bois de tension (Andersson-Gunneras *et al.*, 2006), soulignant ainsi l'importance de ce gène dans l'organisation des microtubules, des microfibrilles de cellulose et dans l'élongation des fibres de bois de tension. En effet, le mutant « *fragile fiber* » *fra2* allélique au gène *AtKTNI* est déficient dans la solidité

## Introduction Générale

des tiges d'inflorescence d'*Arabidopsis* (Burk *et al.*, 2001). De plus, la mutation *fra2* cause une réduction dramatique de la longueur des fibres et de l'épaisseur des parois primaires et secondaires, ainsi qu'une diminution du contenu en cellulose et hémicellulose. Enfin, corrélée avec la désorganisation des microtubules corticaux, cette mutation altère l'orientation normale des microfibrilles de cellulose dans les cellules en expansion (Burk & Ye, 2002).

Dans le bois de tension, le cytosquelette d'actine pourrait avoir une fonction dans la perception et la transduction des contraintes mécaniques et dans le transport des précurseurs pariétaux pour la synthèse de la couche G. D'ailleurs, les transcrits des gènes *ACT2* et *ACT7* codant pour les monomères d'actine sont plus abondants dans le bois de tension que dans le bois normal (Andersson-Gunneras *et al.*, 2006). Curieusement, l'expression de gènes codant pour des chaînes lourdes de myosine est, soit induite, soit réprimée dans le bois de tension. La myosine serait potentiellement importante dans la perception ou la transduction du signal mécanique (Baluska *et al.*, 2003). Dans les cellules du bois, la myosine de type VIII est localisée au niveau de la plaque cellulaire lors de la cytokinèse des initiales fusiformes, et au niveau de toutes les ponctuations entre les fibres, vaisseaux et rayons (Chaffey & Barlow, 2002). La myosine est particulièrement abondante au niveau des plasmodesmes entre les cellules de rayons, et est également trouvée à la périphérie des perforations simples aux extrémités des éléments de vaisseaux et à la périphérie des plaques criblées du phloème suggérant un système de contraction actine/myosine à ces endroits (Chaffey & Barlow, 2002). Toutefois, l'étude de la localisation de la myosine dans les fibres de bois tendu n'a pas encore été réalisée. Dans l'ensemble, l'expression de gènes homologues à d'autres protéines se fixant à l'actine ou « actin binding protein » (ABP) semble non régulée à légèrement réprimée dans le bois de tension de peuplier (Andersson-Gunneras *et al.*, 2006). Néanmoins, l'annotation de ces gènes est basée sur les homologues les plus proches d'*Arabidopsis thaliana* et la caractérisation des familles correspondantes reste à faire chez le peuplier. Notons toutefois, une augmentation de l'expression dans le bois de tension de l'homologue du gène *villine 2* (*VLN2*) d'*Arabidopsis*. Les protéines villines et fimbrines (FIM) pontent et empaquettent les microfilaments d'actine, et il a été montré que la protéine VLN1 lie l'actine et génère des câbles d'actine résistants à la dépolymérisation. VLN1 protège les microfilaments d'actine contre la dépolymérisation par les « actin-depolymerizing factors » (ADF) (Huang *et al.*, 2005). De manière très intéressante, les gènes codant pour des homologues à certaines sous unités « actin related protein » (ARP) du complexe ARP2/3 sont régulés de manière différentielle dans le bois de tension, avec soit une forte répression, soit une forte induction dans le bois de tension. Les transcrits homologues à la sous unité ARPC3 du complexe

## Introduction Générale

ARP2/3 sont les transcrits s'accumulant le plus fortement dans le bois de tension, devant même les FLAs et l'ACC oxydase avec un rapport d'expression entre le bois tendu et le bois normal supérieur à 20 (Andersson-Gunneras *et al.*, 2006). Le complexe ARP2/3 comprenant sept protéines dont ARPC3 est extrêmement conservé chez tous les eucaryotes et active la nucléation et le branchement des microfilaments d'actine. Bien que le gène *ARPC3* et son mutant n'aient pas encore été caractérisés chez les plantes (Szymanski, 2005), des expériences *in vitro* montrent que la perte de la protéine ARPC3 induit une diminution d'un facteur 12 de l'activité du complexe ARP2/3 humain (Gournier *et al.*, 2001). Chez les plantes, les protéines ARP2, ARP3 et ARPC5 de ce complexe sont importantes dans l'établissement de la forme des trichomes, des cellules épidermiques, des cellules de l'hypocotyle et des poils racinaires (Le *et al.*, 2003 ; Li *et al.*, 2003 ; Mathur *et al.*, 2003a ; Mathur *et al.*, 2003b). Les mutants respectifs « *wurm* » « *distorted 1* » et « *crooked* » ont permis de comprendre que ces gènes sont importants dans l'établissement de domaines sub-cellulaires corticaux d'actine filamenteuse fine participant à l'expansion cellulaire localisée de ces cellules. Enfin, l'expression de certains gènes codant pour des protéines LIM est induite dans le bois de tension (Déjardin *et al.*, 2004 ; Andersson-Gunneras *et al.*, 2006). Comme nous le verrons plus tard, les protéines à domaine LIM sont capables de lier les microfilaments d'actine et favorisent la formation de câbles d'actine (Thomas *et al.*, 2006), suggérant une implication de ces protéines dans la perception ou la transduction de la perturbation mécanique dans le bois de tension.

Dans le bois de compression où la forme et la taille des trachéides sont modifiées, des analyses protéomiques ont aussi montré que l'expression de certaines protéines du cytosquelette est dérégulée (Gion *et al.*, 2005). En effet, la profiline et certaines isoformes d'actine et de  $\beta$ -tubulines sont plus abondantes dans le bois opposé que dans le bois de compression. La profiline est une protéine très conservée se liant aux monomères d'actine et impliquée dans l'organisation du cytosquelette chez les eucaryotes. Comme les filaments d'actine sont connus pour réguler l'élongation cellulaire et la forme des cellules, et que la profiline joue un rôle important dans l'établissement de l'élongation et de la croissance polarisée des cellules surtout des poils racinaires (Ramachandran *et al.*, 2000), la sous régulation de ces protéines dans le bois de compression expliquerait en partie la forme arrondie et la taille plus petite des trachéides dans ce bois (Gion *et al.*, 2005). L'ensemble de ces résultats montre que les protéines associées au cytosquelette ont un rôle dans la formation du bois de réaction.

### 3.9.4. Les gènes impliqués dans la synthèse de la lignine

De par la présence d'une couche G enrichie en cellulose, le bois de tension contient beaucoup moins de lignine par rapport aux bois opposé et normal. Ainsi le flux de carbone est préférentiellement redirigé vers la biosynthèse de cellulose par rapport à la lignine. En effet, il est important de constater dans un premier temps que l'expression de gènes impliqués dans la voie des shikimates, la voie de biosynthèse des acides aminés aromatiques conduisant à la synthèse de la phénylalanine, est diminuée dans le bois de tension (Andersson-Gunneras *et al.*, 2006). Ainsi les gènes *DAHP synthase*, *DHQ déhydratases*, *shikimate kinase* et *préphénate déhydratase* présentent une quantité de transcrit fortement réduite dans le bois de tension. Ainsi, lors de la formation du bois de tension, l'arbre favoriserait la diminution de la quantité de lignine par la diminution de la quantité des précurseurs entrant dans la voie de phénylpropanoïdes telle que la phénylalanine.

Concernant les enzymes de la voie des phénylpropanoïdes, des études transcriptomiques récentes confirment en effet une tendance générale à la répression de l'expression des gènes impliqués dans la synthèse de la lignine dans le bois de tension (Pilate *et al.*, 2004a ; Andersson-Gunneras *et al.*, 2006). Premièrement, l'analyse de la distribution des ESTs de l'INRA d'Orléans et des études préliminaires de microarray chez *P. tremula* × *P. alba* montrent que l'expression des gènes *COMT1*, *CCoAOMT1*, *4CL3* et de trois homologues au gène *PAL* est nettement réprimée dans le bois de tension, alors qu'en condition normale et dans le bois opposé, ces gènes sont fortement exprimés dans le xylème en développement (Pilate *et al.*, 2004a) (Leplé, résultats non publiés). Chez le peuplier tremble (*P. tremula*), les résultats sont légèrement différents de par la comparaison de l'expression des gènes entre le bois tendu d'arbre incliné et le bois normal d'arbre droit (Andersson-Gunneras *et al.*, 2006). Les gènes *COMT1*, *PAL1*, *CcOAOMT2*, *F5H*, et *CCR* ont une expression fortement réduite dans le bois de tension, tandis que l'expression des gènes *COMT2*, *C3H*, *4CL3*, *CAD*, *HCT1* et 2 et *CCoAOMT1* est moyennement à faiblement réprimée dans le bois de tension. A l'inverse, on peut noter l'induction de l'expression du gène *CCoAOMT3* dans le bois de tension. Toutefois, les gènes impliqués dans la lignification, excepté le gène *CAD*, font partie de petites familles multigéniques, et les annotations de ces gènes sont à préciser notamment pour les gènes *F5H*, *C3H*, et *CCR*. Les deux gènes *CCoAOMT1* et 2 de peuplier sont tous les deux exprimés dans les cellules lignifiées du xylème, et préférentiellement dans les cellules des rayons et les éléments de vaisseaux en condition de croissance normale (Chen *et al.*, 2000). Ces deux gènes sont induits dans les fibres de xylème lors de l'inclinaison des tiges, et

le profil d'expression spécifique au niveau des cellules de rayon et des vaisseaux est perdu (Chen *et al.*, 2000). Toutefois, cette induction de l'expression des gènes *CCoAOMT1* et 2 serait plutôt liée au stress mécanique ou la blessure plutôt qu'à la formation de bois de tension puisque celui-ci n'est pas encore visible dans ce traitement. Enfin, l'expression de gènes codant pour des homologues aux peroxydases 12 et 64 (PER12 et PER64) d'*Arabidopsis* est fortement induite dans le bois de tension, confortant les résultats ultérieurs montrant une augmentation de l'activité des peroxydases pariétales dans le bois de tension d'eucalyptus (Aoyama *et al.*, 2001 ; Andersson-Gunneras *et al.*, 2006).

### 3.9.5. Régulation de l'expression des gènes dans le bois de tension, les facteurs de transcription

L'ensemble de ces données montre un important niveau de régulation transcriptionnel de gènes impliqués dans divers processus physiologiques lors de la formation du bois de tension. Cette régulation doit probablement être en grande partie orchestrée par des facteurs de transcription. Les premiers résultats d'expression par microarray nous offrent des pistes sur les facteurs de transcription potentiellement impliqués dans la formation du bois de tension au stade tardif de sa formation (Andersson-Gunneras *et al.*, 2006). Avant cela, une première étude a rapporté l'identification d'un facteur de transcription MYB fortement exprimé dans les tissus vasculaires du bois au stade de formation de la paroi secondaire et potentiellement impliqué dans la formation du bois de tension chez *Populus tremula* × *P tremuloides* (Karpinska *et al.*, 2004). L'expression du gène *PttMYB21a* augmente transitoirement dans le xylème durant les premières heures suivant l'inclinaison des arbres et après 12 heures d'inclinaison le niveau d'expression du gène est comparable à celui des arbres témoins non inclinés. Au vu de ces résultats ce gène MYB ne semble pas directement impliqué dans la formation de la couche G, et l'induction de son expression pourrait être le résultat d'une réponse à la blessure ou à la gravité (Karpinska *et al.*, 2004). Toutefois, l'analyse microarray ultérieure montre que ce gène est légèrement induit dans le bois de tension après 11 jours d'inclinaison (Andersson-Gunneras *et al.*, 2006). Cette analyse a aussi identifié une série de gènes codant pour des facteurs de transcription putatifs et dérégulés lors de la formation du bois de tension. Ces gènes appartiennent aux familles de facteurs de transcription WRKY, bHLH, à homéodomaine, et à doigt de zinc de type LIM, GATA, et RING. Toutefois, concernant la fonction des protéines à domaine RING il n'a pas encore été prouvé expérimentalement qu'elles agissent comme des facteurs de transcription, bien qu'elles

## Introduction Générale

puissent avoir une fonction dans la régulation de la transcription des gènes (Takatsuji, 1998). Ce domaine est plutôt considéré comme un domaine d'interaction protéine/protéine, et les protéines connues possédant ce domaine ont aussi une fonction liée à la dégradation des protéines par le protéasome (Takatsuji, 1998 ; Kosarev *et al.*, 2002). Bien que leurs liaisons à l'ADN ne puisse être exclue, leur classification comme facteurs de transcription semble actuellement abusive chez les plantes.

Enfin, l'analyse de la distribution des ESTs dans les banques d'EST de l'INRA d'Orléans montre qu'un cluster d'ESTs homologue au gène *SF3* de tournesol codant pour une protéine à domaine LIM est surreprésenté dans les banques de xylème de bois de tension et de xylème mature (Baltz *et al.*, 1992a ; Déjardin *et al.*, 2004). De la même manière, l'analyse de l'expression des gènes par microarray entre le bois de tension et le bois normal chez *Populus tremula* × *P. tremuloides* a révélé qu'un gène homologue au facteur de transcription LIM est induit d'un facteur 1,85 dans le bois tendu (Andersson-Gunneras *et al.*, 2006). Comme nous le verrons plus tard, parmi les ESTs codant pour des facteurs de transcriptions, les ESTs codant pour des homologues aux protéines à domaine LIM sont les plus abondantes dans les banques d'ADNc de l'INRA d'Orléans. C'est pourquoi, nous avons choisi d'étudier ces protéines à domaines LIM chez le peuplier car elles apparaissent potentiellement impliquées dans le développement des tissus vasculaires du tronc et plus particulièrement dans la formation du bois de tension.

## 4 Les protéines à domaine LIM

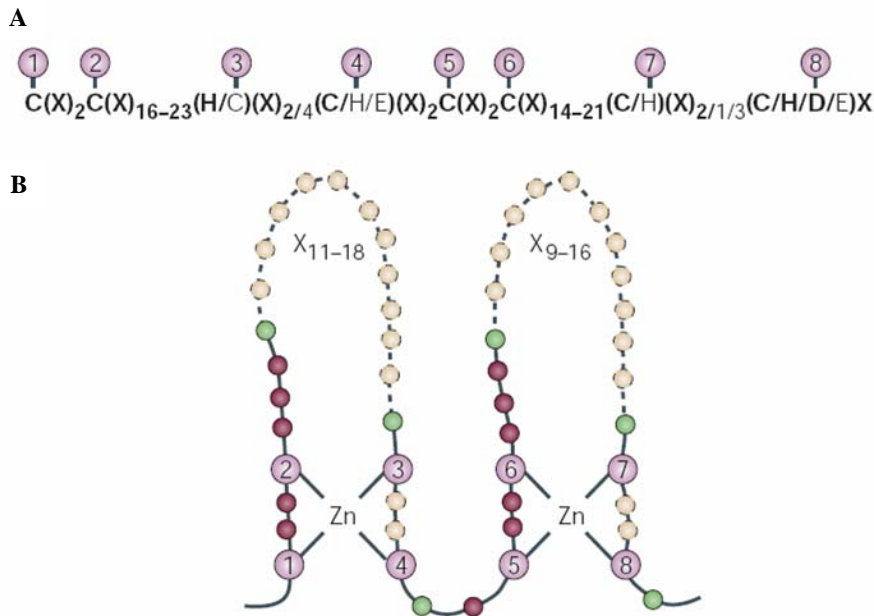
### 4.1. Les protéines à domaine LIM chez les animaux

Les protéines à domaine LIM eucaryotes jouent un rôle important dans les processus biologiques fondamentaux tels que l'organisation du cytosquelette, la transcription, la détermination des lignées cellulaires, le développement des organes et la régulation de fonctions pathologiques comme l'oncogenèse (Bach, 2000). Le nom du domaine LIM dérive des initiales des trois premiers facteurs de transcription découverts : LIN-11 de *Caenorabibditis elegans* (Freyd *et al.*, 1990), ISL-1 du rat (Karlsson *et al.*, 1990) et MEC-3 de *C. elegans* (Way & Chalfie, 1988). Ces trois protéines contiennent en plus de leurs deux domaines LIM riches en cystéines, un homéodomaine supplémentaire permettant la fixation à l'ADN. Depuis, une grande variété de protéines contenant un ou plusieurs domaines LIM a été décrite chez les eucaryotes. Toutes les données bibliographiques montrent que les protéines LIM forment une famille complexe et présentent une grande variété de fonctions essentielles dans divers processus biologiques. Les protéines LIM sont connues pour être nucléaires, cytoplasmiques ou faisant la navette entre ces deux compartiments cellulaires. De plus, les protéines à domaine LIM sont capables d'interagir avec elles même, d'autres protéines à domaine LIM ou des protéines non apparentées. Ainsi, le domaine LIM est un module d'interaction protéine/protéine et facilite l'assemblage de complexes multiprotéiques dans la cellule.

#### 4.1.1. Séquence et structure des domaines LIM

Chaque domaine LIM individuel comprend approximativement 55 acides aminés avec huit résidus très conservés, surtout des cystéines et histidines, ayant des positions bien précises. La séquence consensus classique du domaine LIM a été définie comme telle [C-X<sub>2</sub>-C-X<sub>16-23</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>16-21</sub>-C-X<sub>2-3</sub>-C(D,H)] (où X représente n'importe quel acide aminé) (Schmeichel & Beckerle, 1994). Toutefois, des variations dans cette séquence consensus ont été constatées avec l'analyse de l'ensemble des protéines LIM humaines et d'autres espèces (Fig. 26). Le nombre et l'espace conservés des huit résidus cystéines et histidines définissent des motifs de liaison aux métaux au sein de ces domaines LIM, et il a

été montré que ce domaine lie préférentiellement le zinc par rapport au fer (Michelsen *et al.*, 1993). Le domaine LIM coordonne donc deux atomes de zinc ce qui définit la topologie de deux motifs à doigt de zinc répétés en tandem. Deux résidus conservés séparent les deux doigts de zinc, et l'espacement entre ces doigts de zinc semble être essentiel à leurs fonctions (Michelsen *et al.*, 1994 ; Schmeichel & Beckerle, 1997). De plus, il y a plusieurs résidus aliphatiques ou volumineux modérément conservés qui sont situés à des positions bien définies (Fig. 26). Le reste de la séquence du domaine LIM est très variable et confère aux protéines LIM leurs spécificités fonctionnelles. Les acides aminés conservés facilitent la formation d'une structure cœur et les acides aminés variables permettent une fixation de haute affinité avec de nombreuses protéines partenaires ayant des structures et des fonctions diverses.



**Figure 26.** Séquence consensus et topologie du domaine LIM (Kadmas et Beckerle, 2004)

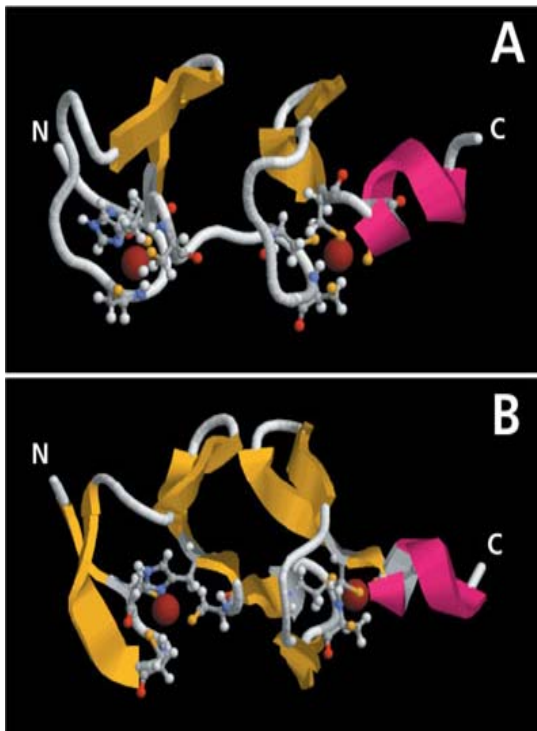
**A.** Espacement et identité des huit résidus liant le zinc, basé sur l'analyse de 135 séquences LIM humaines. Les acides aminés rarement observés (inférieur à 10 % des cas) ne sont pas indiqués en gras. X représente n'importe quel acide aminé.

**B.** Topologie de la structure du domaine LIM et de la coordination au zinc. Les ronds violets indiquent les résidus se liant au zinc. Les résidus aliphatiques ou volumineux semi conservés sont représentés par des ronds verts, et les résidus non conservés avec un espacement invariant sont figurés par des ronds magentas. Les lignes hachurées et les ronds jaunes indiquent le nombre variable de résidus (X) au sein des boucles des doigts de zinc.

La détermination par spectroscopie RMN de la structure tridimensionnelle du domaine LIM C-terminal de la protéine « *Cystein Rich Protein 1* » (CRP1) montre que les deux modules contenant un atome de zinc (CCHC et CCCC) comprennent deux feuilletts  $\beta$  antiparallèles arrangés de manière orthogonale, et le deuxième module (CCCC) se termine par



une hélice  $\alpha$  (Fig. 27) (Perez-Alvarado *et al.*, 1994). Les deux doigt de zinc constituant le domaine LIM représentent des entités structurales séparées qui sont tenues ensemble dans une configuration particulière par un réseau d'interactions hydrophobes. La coordination conservée tétraédrique du zinc établit la structure secondaire et le repliement tertiaire du domaine LIM et au niveau des boucles, plusieurs résidus hydrophobes importants aident à l'emballage des deux doigts de zinc. Des études RMN sur d'autres protéines LIM telles que les protéines LMO ont montré la conservation du repliement des domaines LIM (Kadrmas & Beckerle, 2004). Le repliement tertiaire des domaines LIM dans le contexte des protéines CRP1 et CRP2 entières est identique à la structure des domaines LIM isolés. Dans ces protéines, il n'y a pas de préférence dans l'organisation spatiale des deux domaines LIM supportant la notion que les deux domaines LIM sont des structures et des modules fonctionnels indépendants et que ces protéines servent d'adaptateurs pour la formation de complexes macromoléculaires.



**Figure 27.** Structure des domaines LIM individuels de la protéine CRP2 (Weiskirchen et Günther, 2003).

Représentation schématique des structure RMN des domaines LIM1 (A) et LIM2 (B) de la protéine CRP2 de caille incluant les atomes des chaînes latérales des résidus cystéines et histidines coordonnant le zinc. Dans ce modèle, les feuillets  $\beta$  sont indiqués en jaune/orange, les hélices  $\alpha$  sont de couleur magenta et les atomes de zinc sont représentés par des sphères rouges.

#### 4.1.2. Les domaines LIM sont des modules d'interaction protéine/protéine et peuvent fixer l'ADN

Les études structurales par RMN montrent que le deuxième module à doigt de zinc du domaine LIM est structurellement similaire au domaine à doigt de zinc de type GATA (Perez-Alvarado *et al.*, 1994), capable de se lier à l'ADN. Bien que ce deuxième doigt de zinc adopte

un repliement théoriquement compatible avec une fixation aux acides nucléiques, le domaine LIM est plutôt impliqué dans des interactions protéine/protéine spécifiques et participerait à la formation de complexes multiprotéiques (Schmeichel & Beckerle, 1994 ; Dawid *et al.*, 1998). Jusqu'à présent, chez les animaux, il n'y a pas d'évidences convaincantes montrant une interaction directe entre les domaines LIM et l'ADN *in vivo*. Ceci est d'autant plus intrigant que beaucoup de protéines à domaine LIM participent au contrôle transcriptionnel dans le noyau. Toutefois, la protéine Hic5 de la famille des paxillines est la seule protéine à domaine LIM de vertébrés connue pour être capable de se lier à l'ADN *in vitro* (Nishiya *et al.*, 1998). La protéine Hic5 se fixe à l'ADN par l'intermédiaire de ces quatre domaines LIM en C-terminal de la protéine et chacun des domaines LIM contribue à la fixation à l'ADN de manière distincte. De plus, chez les plantes il a été montré que les protéines SF3 (HaPLIM1) et NtLIM1 peuvent se fixer à l'ADN (Baltz *et al.*, 1996 ; Kawaoka *et al.*, 2000). Chez les animaux, les domaines LIM sont plus connus comme participant à des interactions protéine/protéine, mais une séquence consensus ou un élément structural capable de se fixer aux domaines LIM est encore à déterminer (Kadrmas & Beckerle, 2004). Le domaine LIM semble fournir un cadre structural stable sur lequel plusieurs sortes de spécificités de liaison peuvent être superposées. Un petit motif peptidique linéaire présentant une haute affinité pour les domaines LIM est actuellement peu probable, et les spécificités de liaison seraient plutôt dispersées au sein des protéines interagissant avec les domaines LIM. Dans beaucoup de cas, le domaine LIM seul est suffisant pour la fixation spécifique d'une protéine partenaire, mais il a été montré que les séquences adjacentes au domaine LIM peuvent intervenir dans la liaison (Arber & Caroni, 1996). De plus, un domaine LIM peut fixer simultanément plusieurs protéines, et des domaines LIM en tandem peuvent aussi fixer un partenaire unique (Schmeichel & Beckerle, 1998 ; Kadrmas & Beckerle, 2004).

### 4.1.3. Les protéines contenant des domaines LIM chez les eucaryotes

Les protéines à domaines LIM sont rencontrées dans une grande variété d'organismes eucaryotes, et tous les génomes séquencés d'eucaryotes codent pour des protéines contenant des domaines LIM. Par contre, aucune protéine à domaine LIM n'a été identifiée chez les procaryotes. Chez l'homme, il existerait au moins 58 gènes codant pour des protéines contenant au moins un domaine LIM (Kadrmas & Beckerle, 2004). Les protéines LIM humaines contiennent un à cinq domaines LIM conservés. Avec 135 domaines LIM

## Introduction Générale

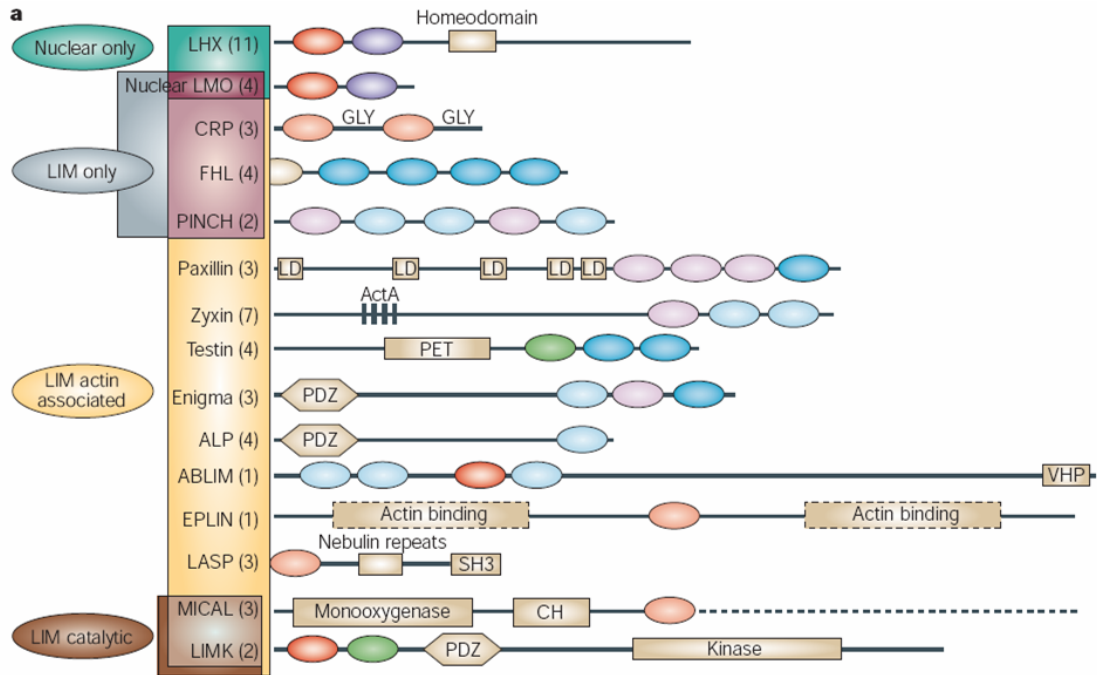
répertoriés chez l'homme, l'abondance de ce domaine dans les protéines est similaire à celui d'autres domaines d'interaction protéine/protéine comme les domaines « *Src homology-2* ou *3* » (SH2 et SH3). Les invertébrés comme la drosophile (*Drosophila melanogaster*) ou *C. elegans* possèdent à peu près les mêmes familles de protéines LIM que les vertébrés, mais le nombre de protéines LIM au sein de chaque famille est plus réduit. Les génomes de levures et de champignons codent pour un petit nombre de protéines à domaine(s) LIM. Par exemple, la levure ne contient que trois protéines à domaine LIM.

Les protéines à domaine LIM humaines qui ont été caractérisées au niveau moléculaire, et les différentes familles correspondantes sont récapitulées dans la figure 28. Les protéines LIM peuvent être composées exclusivement de domaines LIM, ou bien les domaines LIM positionnés en N-terminal ou C-terminal des protéines peuvent être associés à un autre domaine comme des homéodomains, des domaines se liant au cytosquelette, des domaines catalytiques ou d'autres domaines d'interaction protéine/protéine tels que les domaines SH3, PDZ ou le motif LD (Fig. 28). Basé sur la structure et l'organisation des domaines LIM, la localisation subcellulaire et la fonction des protéines LIM, les familles de protéines LIM ont été classées en quatre groupes majeurs qui se recoupent entre eux (Bach, 2000 ; Kadrmas & Beckerle, 2004) : Ce sont les groupes « *Nuclear only* », « *LIM only* », « *LIM actin associated* » et « *LIM catalytic* ». Les similarités de séquences en acides aminés entre les domaines LIM définissant ces quatre catégories sont illustrées sur la figure 28 par des ovales de différentes couleurs. Les différents membres de chaque famille de protéines LIM portent souvent le même arrangement de domaines LIM, indiquant probablement une expansion de ces familles à partir d'une protéine ancestrale commune. Les domaines LIM1 (en rouge) ou LIM2 (en violet) des protéines nucléaires LHX et LMO ont une forte similarité de séquence entre eux et se distinguent globalement des autres domaines LIM appartenant aux protéines cytoplasmiques (Kadrmas & Beckerle, 2004). Au sein de chaque famille de protéines CRP et FHL, les multiples domaines LIM sont fortement apparentés entre eux, suggérant qu'ils sont apparus après un événement récent de duplication d'un domaine LIM ancestral (Fig. 28). Enfin, pour chacune des familles de protéines zyxine, paxilline, enigma, ALP et EPLIN, les domaines LIM composant la protéine sont en général divergents entre eux et ont ainsi pu être assemblés suite à la fusion de domaines LIM structurellement différents.

Ainsi, les génomes des animaux contiennent un grand nombre de gènes codant pour des protéines contenant un nombre variable de domaine LIM, associé ou non avec d'autres domaines. Il existe donc pour les protéines LIM animales une grande plasticité structurelle et en conséquence une importante diversité fonctionnelle. Pour plus d'informations concernant

## Introduction Générale

les fonctions des protéines appartenant à ces familles LIM, je vous invite à la lecture des revues bibliographiques suivantes (Bach, 2000 ; Kadrmas & Beckerle, 2004). A l'inverse, les génomes de plantes semblent coder pour un nombre limité de protéines à domaine LIM, et celles découvertes jusqu'à présent sont toutes apparentées à la famille de protéines riches en cystéine (CRP) animale (Eliasson *et al.*, 2000). Par ailleurs, les fonctions des protéines zyxine et Hic5 seront d'abord décrites ci-dessous car elles interagissent avec les protéines CRP.



**Figure 28.** Représentation schématique de la structure protéique des différentes familles de protéines à domaines LIM les mieux caractérisées chez l'Homme (Kadrmas et Beckerle, 2004)

Le nombre de gènes connus dans chaque famille est indiqué entre parenthèses. Les boîtes de différentes couleurs représentent les différentes catégories de protéines LIM communément définies. Les domaines LIM sont figurés par des ovales de différentes couleurs et ont été groupés en fonction de leurs similarités de séquence en acides aminés. Les autres domaines hétérologues comme les motifs LD, les domaines de liaison à l'actine, monoxygénase, SH3, et homeodomaine sont marqués par des rectangles bruns. Les rectangles hachurés indiquent les domaines dont les limites sont mal définies. Les lignes hachurées montrent que l'échelle n'est pas respectée.

ABLIM : actin-binding LIM protein ; ALP :  $\alpha$ -actinin-associated LIM protein ; CH : calponin homology ; CRP : Cysteine-rich protein ; EPLIN : epithelial protein lost in neoplasm ; FHL : Four-and-a-half LIM ; GLY : glycine-rich region (GRR) ; LASP : LIM and SH3 protein ; LD : motif LD ayant la séquence consensus [LDXLLXXL] ; LHX : LIM-homeodomain protein ; LMK : LIM kinase ; LMO : LIM only ; MICAL : Molecule interacting with CASL protein-1 ; PDZ : postsynaptic density-95 (PSD95), Disc large, zona occludens-1 (ZO-1) ; PET : prickle spinas and testin ; PINCH Particularly interesting new cysteine and histidine-rich protein ; SH3, Src homology-3 ; VHP : villin head piece.

### 4.1.4. Les protéines zyxine et Hic5 sont associées aux adhésions focales et fibres de stress d'actine

Beaucoup de résultats montrent l'interaction de protéines à domaine LIM (dont CRP1 et CRP2) avec les fibres de stress d'actine intervenant dans la contraction des cellules animales. Ces fibres particulières sont des câbles de filaments d'actine contractiles et sont formées dans la cellule après attachements des filaments d'actine sur des complexes de protéines situés au niveau des adhésions focales. Les adhésions focales sont les sites où les cellules adhèrent à la matrice extracellulaire par l'intermédiaire des intégrines. Elles ont un rôle structural de liaison entre la matrice extracellulaire et le cytoplasme, mais aussi dans la transduction du signal et dans la diffusion et la migration cellulaire. Ces complexes de protéines situés au niveau des adhésions focales comprennent entre autre les intégrines, le complexe Arp2/3, la zyxine, l' $\alpha$ -actinine, Hic5 et les CRPs. La formation des fibres de stress d'actine et des plaques d'adhésion est induite par la contraction/tension des cellules.

La zyxine est une protéine contenant trois domaines LIM en C-terminal, une région N-terminale riche en proline et un signal nucléaire d'exportation (Sadler *et al.*, 1992 ; Wang & Gilmore, 2003). C'est une protéine ayant de multiples partenaires dans la cellule dont des protéines interagissant avec le cytosquelette ou la membrane plasmique, des molécules signal et des facteurs de transcription. Notamment, la zyxine interagit par ses domaines LIM avec les protéines CRP (Sadler *et al.*, 1992) et avec l' $\alpha$ -actinine par sa région riche en proline (Reinhard *et al.*, 1999). La zyxine présente une double localisation nucléaire et cytoplasmique. La phase nucléaire de la zyxine est transitoire et dans le cytoplasme elle est située le long des fibres de stress d'actine et au niveau des adhésions focales (Nix & Beckerle, 1997). De façon intéressante, la zyxine est capable de se diriger vers le noyau dans les cellules musculaires lisses vasculaires stimulées mécaniquement (Cattaruzza *et al.*, 2004). De plus, dans un contexte cellulaire différent, la zyxine est aussi capable sous l'influence d'un stress mécanique d'étirement cyclique des cellules, de se mobiliser des adhésions focales vers les fibres de stress d'actine qui s'épaississent (Yoshigi *et al.*, 2005). La zyxine est donc une protéine faisant la navette entre les adhésions focales, les fibres de stress d'actine et le noyau.

La protéine Hic5 appartient à la famille des paxillines. La paxilline et Hic5 sont des protéines localisées au niveau des adhésions focales, et interagissent avec de nombreux partenaires protéiques par l'intermédiaire des quatre domaines LIM en C-terminal, et des motifs LD en N-terminal (Kadmas & Beckerle, 2004). Contrairement à la paxilline ayant une expression ubiquitaire, la protéine Hic5 est exprimée seulement dans les cellules musculaires

lisses de souris (Kim-Kaneyama *et al.*, 2005). De façon intéressante, les domaines LIM2 et LIM3 en C-terminal sont responsables de la localisation de la protéine Hic5 sur les adhésions focales en condition normale (non étirée) et le long des fibres de stress d'actine quand un étirement mécanique est appliqué aux cellules. De plus Hic5 est capable d'interagir avec la protéine CRP2 par ses domaines LIM en C-terminal. Enfin, la protéine Hic5 a un effet suppressif sur la capacité contractile des cellules entourées d'un gel de collagène tridimensionnel, et cet effet est aboli quand les domaines LIM2 et LIM3 sont mutés. Ces résultats suggèrent qu'en réponse à un étirement mécanique cyclique, la protéine Hic5 est relocalisée vers les fibres de stress d'actine et régule négativement la force contractile des cellules (Kim-Kaneyama *et al.*, 2005). Il apparaît ainsi que le complexe Hic5/CRP2/zyxine pourrait réguler l'organisation du cytosquelette d'actine lors de stress mécanique.

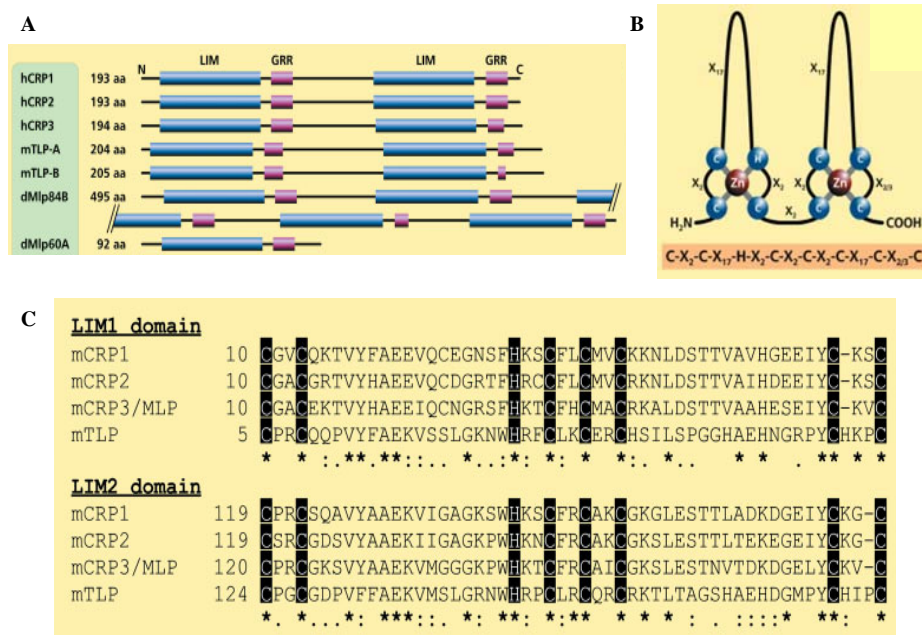
### 4.2. La famille de protéines riches en cystéines (CRP) chez les eucaryotes

#### 4.2.1. Structure des protéines CRP

Au sein de la famille des protéines riches en cystéines (CRP), il existe quatre membres chez l'homme et les animaux : CRP1, CRP2, CRP3/MLP pour « *muscle LIM protein* » et TLP pour « *thymus LIM protein* ». Les trois protéines CRP présentent une forte identité de séquence (supérieure à 65 %) et sont composées uniquement de deux domaines LIM séparés par un long espaceur de 56 à 59 acides aminés (Fig. 29A) (Weiskirchen & Günther, 2003). La protéine TLP est plus divergente avec en moyenne 30 % d'identité en acides aminés avec les protéines CRP (Kirchner *et al.*, 2001). Contrairement, aux autres protéines possédant de multiples domaines LIM séparés entre eux par 8 à 10 acides aminés, la région flexible liant les deux domaines LIM des protéines CRP est nettement plus longue d'environ 50 acides aminés, et caractérise cette famille de protéine. La taille de ces protéines (environ 200 acides aminés), ainsi que la taille des domaines LIM et l'espacement entre chaque domaine LIM sont très conservés. Les domaines LIM des protéines CRP mesurent de 52 à 53 acides aminés et présentent la séquence consensus suivante [C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2-3</sub>-C] (Fig. 29B). Nous pouvons noter que la protéine TLP diffère dans la composition de ces deux domaines LIM par l'addition d'un résidu histidine supplémentaire suivant la troisième cystéine du second doigt de zinc (Fig. 29C). Chaque domaine LIM est suivi d'une région riche en glycine (GRR) marquant la distinction entre les protéines CRP et les autres protéines

## Introduction Générale

LIM-only. De plus, les protéines CRP1 à 3 possèdent une séquence putative d'adressage nucléaire [K/R-K/R-Y-G-P-K] au niveau des acides aminés 64 à 69 et qui chevauche partiellement la région GRR suivant le premier domaine LIM. Par ailleurs, chez la drosophile (*Drosophila melanogaster*) deux protéines CRP atypiques Mlp60A et Mlp84B pour « *Muscle lim proteins* » ont été découvertes (Stronach *et al.*, 1996). Mlp60A ne possède qu'un domaine LIM alors que Mlp84B en contient cinq (Fig. 29A). Ces domaines LIM très similaires à ceux des protéines CRP, sont également suivis de régions GRR et ces deux protéines possèdent le signal putatif d'adressage nucléaire. Enfin, la protéine DdLIM, renommée plus tard DdLIME, est la première protéine à domaine LIM isolée chez l'amibe *Dictyostelium discoideum* (Prassler *et al.*, 1998). Elle comporte un seul domaine LIM suivi d'une région riche en glycines typique du groupe de protéines CRP, mais plus étendue. Plus récemment, les gènes *DdLIMC* et *D* ont été caractérisés et codent aussi pour des protéines homologues aux protéines CRP animales, sauf que ces protéines ne possèdent pas de région GRR et DdLIMD ne contient qu'un seul domaine LIM en N-terminal (Khurana *et al.*, 2002).



**Figure 29.** Structure des protéines « Cystein Rich Protein » (CRP) animales et de leurs deux domaines LIM (Weiskirchen & Günther, 2003).

**A.** Représentation schématique de la topologie des protéines CRP de l'Homme (hCRP1, hCRP2, hCRP3/MLP), et CRP-like de la souris (mTLP-A, mTLP-B) et de la drosophile (dMlp84B, dMlp60A). Le nom des protéines et le nombre total d'acides aminés sont donnés à gauche de chaque protéine. Les domaines LIM sont indiqués par des boîtes bleues et les positions des régions riches en glycines (GRR) sont représentées par des boîtes de couleur magenta. **B.** Topologie des deux structures en doigts de zinc (CCHC et CCCC), et séquence consensus des deux domaines LIM des protéines CRP1, CRP2, CRP3/MLP et TLP. **C.** Alignement multiple de séquences (ClustalW) des domaines LIM1 et LIM2 issus des protéines CRP1 à 3 et TLP de souris. Les résidus cystéines et histidines conservés coordonnant les atomes de zinc des deux domaines LIM sont indiqués par un arrière fond noir. Les résidus identiques, similaires ou présentant une faible similarité biochimique sont respectivement indiqués par un astérisque, un double point et un simple point.

4.2.2. Les gènes CRP sont préférentiellement exprimés dans les cellules musculaires

Bien que les protéines CRP soient structurellement très similaires, les gènes codant les protéines CRP présentent des profils d'expression différents (Tableau 6), suggérant que les trois membres de la famille CRP (CRP1, 2 et 3) accomplissent des fonctions similaires dans différents tissus. Les gènes *CRP* sont surtout exprimés dans les cellules musculaires, et les membres de cette famille sont probablement impliqués dans la différenciation terminale du développement des muscles de vertébrés. Le gène *CRP1* de poulet et la protéine correspondante sont surtout exprimés dans les tissus comportant des muscles lisses comme l'estomac, les intestins et le gésier (Crawford *et al.*, 1994 ; Louis *et al.*, 1997). De plus, le niveau d'expression du gène *CRP1* augmente fortement durant la maturation des muscles lisses. La présence de la protéine CRP2 est restreinte aux artères comportant des muscles lisses vasculaires et aux fibroblastes. L'abondance de la protéine CRP2 dans les artères augmente en fonction de l'âge de l'embryon. La protéine CRP3/MLP est plus fortement exprimée dans les tissus comportant des muscles striés comme le muscle cardiaque (Louis *et al.*, 1997). De plus, l'expression du gène *CRP3/MLP* augmente quand les myoblastes commencent à se différencier (Arber *et al.*, 1994). Ainsi, l'expression des gènes *CRP1*, *CRP2* et *CRP3/MLP* et de leurs protéines augmente en fonction du stade de développement des organes. Enfin, le gène et la protéine TLP sont spécifiquement exprimés dans le thymus au niveau d'un sous ensemble de cellules épithéliales corticales (Kirchner *et al.*, 2001).

**Tableau 6.** Expression des protéines « *cystein rich protein* » (CRP) et protéines partenaires interagissant avec elles (Weiskirchen et Günther, 2003).

Protéines	Expression tissulaire préférentielle	Partenaires cytoplasmiques	Partenaires nucléaires
CRP1	Poumons, intestins, artères, estomac, fibroblastes	CRP1, zyxine, $\alpha$ -actinine, actine	GATA6 ?
CRP2	Artères, fibroblastes	zyxine, $\alpha$ -actinine, Hic5, actine	GATA6, GATA4, SRF, Brg1, CRP2BP, PIAS
CRP3/MLP	Coeur, muscles squelettiques	CRP3/MLP, zyxine, $\alpha$ -actinine, $\beta$ I-spectrine, T-cap	Apterous, RBTN1 (LMO1), MyoD, MRF4, myogénine.

CRP2BP : CRP2 binding protein ; LMO1 : LIM only 1 ; MLP : Muscle LIM protein ; MRF4 : muscle regulatory factor 4 ; PIAS : protein inhibitor of activated STAT1 ; RBTN1 : rhombotin-1 ; SRF : serum response factor. D'après (Knoll *et al.*, 2002 ; Chang *et al.*, 2003 ; Weiskirchen & Günther, 2003 ; Grubinger & Gimona, 2004 ; Kim-Kaneyama *et al.*, 2005 ; Tran *et al.*, 2005 ; Chang *et al.*, 2007).



### 4.2.3. Les protéines CRP ont une double localisation nucléaire et cytoplasmique

Une des caractéristiques intéressantes des protéines CRP1, CRP2 et CRP3/MLP est qu'elles présentent une distribution cellulaire bimodale. Les protéines riches en cystéines CRP des vertébrés s'accumulent à la fois dans le noyau et le cytoplasme où elles s'associent avec le cytosquelette d'actine (Arber & Caroni, 1996). Nous avons vu précédemment que les protéines CRP ont un signal d'adressage nucléaire putatif dans la première région GRR. Au contraire, la protéine TLP ne possède pas ce signal d'adressage nucléaire, et est uniquement localisée dans le cytoplasme parfois associée avec l'actine (Kirchner *et al.*, 2001). La protéine de drosophile Mlp60A/DMLP1 est exclusivement nucléaire (Arber & Caroni, 1996). Par ailleurs, la localisation des protéines CRP est influencée par le contexte cellulaire, et semble varier en fonction de l'âge des tissus, notamment entre les tissus embryonnaires et adultes. En effet, la localisation nucléaire ou cytoplasmique des protéines CRP2 et CRP3 dépend de l'état de différenciation des cellules. La protéine CRP2 est localisée dans le noyau dans les cellules proépicardiales, et sa localisation devient à la fois nucléaire et cytoplasmique lorsque ces cellules se différencient en cellules de muscle lisse (Chang *et al.*, 2003). De même, dans les fibroblastes indifférenciés, la protéine CRP3/MLP est exclusivement nucléaire, alors que dans les cellules musculaires différenciées, elle est présente dans le noyau et le cytoplasme (Arber *et al.*, 1994 ; Arber & Caroni, 1996).

### 4.2.4. Dans le noyau, les protéines CRP interagissent avec des facteurs de transcription

Dans le noyau, les protéines CRP sont impliquées dans des processus de différenciation cellulaire en agissant comme des cofacteurs de transcription (Tableau 6) (Kong *et al.*, 1997 ; Chang *et al.*, 2003). Avec les facteurs de transcription GATA6 et SRF (le facteur de réponse au sérum), les protéines CRP1 et CRP2 sont capables d'activer la transcription de gènes spécifiques des muscles lisses. Les protéines CRP2, GATA6 et SRF interagissent ensemble de façon directe et forment un complexe ayant une forte affinité pour l'élément *cis* activateur de fixation au SRF « *Serum Response Element* » (SRE). Ce complexe GATA6/SRF/CRP2 en se fixant à l'élément SRE, active fortement les promoteurs des gènes spécifiques des muscles lisses dans des fibroblastes pluripotents, et pourrait ainsi induire leur conversion en cellules musculaires lisses (Chang *et al.*, 2003). Le domaine LIM N-terminal de la protéine CRP2 s'associe avec la protéine SRF et le domaine LIM C-terminal s'associe avec

## Introduction Générale

la protéine GATA4, ce qui permet le rapprochement de ces protéines. Les deux doigts de zinc pour chaque domaine LIM sont nécessaires à ces deux types d'interactions, et le domaine LIM N-terminal a un rôle fonctionnel important nécessaire à l'activation synergique du promoteur. De plus, une protéine CRP2 mutante dominante négative ne contenant pas le domaine LIM N-terminal et exprimée dans les cellules proépicardiales, bloque l'expression des gènes spécifiques de muscle lisse et la différenciation en cellules musculaires vasculaires (Chang *et al.*, 2003). Une étude récente montre que l'expression de la protéine CRP2 dirigée dans le cœur adulte de souris transgénique conduit à l'expression des gènes spécifiques de cellules musculaires lisse. Dans ces conditions la protéine CRP2 se fixe *in vivo* sur les promoteurs des gènes codant pour la calponine, la  $\gamma$ -actine et la chaîne lourde de myosine de muscles lisses (SM-MHC) de manière indirecte probablement à travers le recrutement du facteur de transcription SRF (Chang *et al.*, 2007). Ainsi les protéines CRP1 et CRP2 pourraient fonctionner comme des cofacteurs de la transcription formant des complexes avec les facteurs de transcription SRF, GATA4 et GATA6 pour promouvoir la régulation de gènes spécifiques des muscles lisses essentiels pour la différenciation des cellules musculaires lisses cardiovasculaires. Par ailleurs, la protéine CRP2 est capable par son domaine LIM N-terminal de recruter et d'interagir directement avec la protéine ATPase Brg1, un cofacteur du complexe SWI/SNF de remodelage de la chromatine (Chang *et al.*, 2007).

Dans le noyau, la protéine CRP3/MLP interagit avec de nombreux partenaires (Tableau 6). En effet, elle se fixe à la protéine Apterous un facteur de transcription LIM à homéodomaine localisé exclusivement dans le noyau, et la protéine LMO rhombotin-1 (RBTN1) une protéine à deux domaines LIM majoritairement localisée dans le noyau (Arber & Caroni, 1996). Par ailleurs, la protéine nucléaire CRP3/MLP favorise la myogenèse en se fixant par son domaine LIM N-terminal à la protéine MyoD, un facteur de transcription de type bHLH spécifique des muscles. Cette fixation induit l'activation de MyoD (Kong *et al.*, 1997). CRP3 interagit aussi avec d'autres protéines de type bHLH comme MRF4 et la myogénine, des « muscle regulatory factor ». Par contre, elle n'interagit pas avec les facteurs de transcription bHLH E47 et E12 non spécifiques des muscles. Bien que la protéine CRP3 ne possède pas de domaine activateur de la transcription, elle pourrait servir de cofacteur au facteur de transcription MyoD en augmentant son interaction avec l'ADN au niveau des éléments régulateurs (Kong *et al.*, 1997). La suppression de l'expression du gène *CRP3* en utilisant la technologie antisens bloque la différenciation des myoblastes en cellules musculaires, et la surexpression du gène dans les myoblastes induit la progression de la différenciation musculaire (Arber *et al.*, 1994). La disruption du gène *CRP3* chez la souris

confirme l'importance de la protéine CRP3 dans le maintien de l'intégralité structurale des cellules musculaires cardiaques et squelettiques (Arber *et al.*, 1997). En effet, les souris CRP3/MLP<sup>-/-</sup> développent après la naissance des cardiomyopathies dilatées avec des hypertrophies du coeur et des arrêts cardiaques conduisant à une mortalité postnatale. Ces symptômes sont le résultat d'une sévère altération de l'organisation de l'architecture cytologique des cardiomyocytes.

### 4.2.5. Les protéines CRP cytoplasmiques sont associées aux filaments d'actine

Dans le cytoplasme, toutes les protéines CRP se lient à des éléments du cytosquelette d'actine (Tableau 6). Il a en effet été montré que les protéines CRP1, 2 et 3 sont toutes capables d'interagir directement avec l' $\alpha$ -actinine une protéine pontant l'actine, et la zyxine une protéine à domaine LIM des plaques d'adhésion focale liant aussi l'actine (Sadler *et al.*, 1992 ; Schmeichel & Beckerle, 1994 ; Louis *et al.*, 1997 ; Pomies *et al.*, 1997). De plus, il a été montré que la protéine CRP1 humaine est capable de former des homodimères (Feuerstein *et al.*, 1994). Cette propriété est portée par les domaines LIM de la protéine, et la présence de doigts de zinc intacts coordonnant deux atomes de zinc est nécessaire et suffisante à la formation de ces homodimères. Cette capacité de dimérisation peut être transférée à d'autres protéines par fusion du domaine LIM. Il a également été montré que la protéine CRP3/MLP est capable de s'associer avec elle-même, et la formation d'homodimère nécessite les deux domaines LIM1 et LIM2 (Arber & Caroni, 1996). Dans le cytoplasme, les protéines zyxine et CRP1 colocalisent et sont détectées par immunofluorescence sur les câbles de filaments d'actine et plus fortement sur les fibres de stress d'actine près de leur attachement sur les plaques d'adhésion focale (Sadler *et al.*, 1992 ; Crawford *et al.*, 1994). L'interaction de la protéine CRP1 avec l' $\alpha$ -actinine est confinée à une séquence de 18 acides aminés localisée dans la région GRR N-terminale suivant le premier domaine LIM (Harper *et al.*, 2000), alors que son interaction avec la zyxine nécessite l'action coordonnée de résidus positionnés dans les deux domaines LIM de la protéine CRP1 (Schmeichel & Beckerle, 1998).

Les scientifiques ont longtemps cru que l'association des protéines CRP avec l'actine était indirecte plutôt que directe. Notamment, les protéines CRP ne contiennent aucun motif de liaison à l'actine. Toutefois chez *Dictyostelium*, les protéines DdLIMC et D interagissent directement avec l'actine *in vitro* (Khurana *et al.*, 2002). De plus, il a été montré récemment par des expériences de cosédimentation que les protéines CRP1 et CRP2 sont capables

## Introduction Générale

d'interagir directement avec l'actine filamenteuse (F-actine) (Grubinger & Gimona, 2004 ; Tran *et al.*, 2005). De plus, la protéine CRP2 fusionnée à la GFP est associée avec les fibres de stress d'actine et cette association est indépendante de la localisation de l' $\alpha$ -actinine et la zyxine dans la cellule (Grubinger & Gimona, 2004). Quand la protéine CRP1 est exprimée dans les fibroblastes embryonnaires de rat, les filaments d'actine sont plus épais et l'empaquetage « *bundling* » des filaments d'actine est ainsi augmenté (Tran *et al.*, 2005). La protéine CRP1 est en effet capable d'empaqueter directement et indépendamment les filaments d'actine. Bien que la protéine CRP1 n'affecte pas l'activité d'empaquetage de la protéine  $\alpha$ -actinine, elle stabilise l'interaction de l' $\alpha$ -actinine avec les câbles d'actine. En utilisant la technique de transfert d'énergie entre molécules fluorescentes (FRET), il a été montré l'existence de deux populations de protéines CRP1 localisées le long des fibres de stress d'actine : l'une associée à l'actine à travers l'interaction avec l' $\alpha$ -actinine, et l'autre directement liée aux filaments d'actine (Tran *et al.*, 2005). Ainsi, les auteurs suggèrent un rôle de CRP1 dans la régulation de l'empaquetage des filaments d'actine par le pontage direct des filaments d'actine et le renforcement de l'interaction entre les câbles d'actine et l' $\alpha$ -actinine.

De façon intéressante, la localisation cellulaire de la protéine CRP2 varie de la même manière que les protéines Hic5 et zyxine en fonction de l'application d'un stress mécanique d'étirement (Kim-Kaneyama *et al.*, 2005 ; Yoshigi *et al.*, 2005). Sous l'effet de l'étirement cyclique des cellules, sa localisation au niveau des fibres de stress d'actine est nettement plus marquée. Au contraire, la protéine CRP1 est distribuée dans le cytoplasme et le noyau, mais pas sur les fibres de stress d'actine lors de l'étirement des cellules. La protéine CRP2 interagit *in vitro* et *in vivo* avec la protéine Hic5. De plus, sous l'effet de l'étirement mécanique des cellules, Hic5 et CRP2 colocalisent et se déplacent vers les fibres de stress d'actine. Enfin, l'expression de la protéine CRP2 réduit la capacité contractile des cellules entourées d'un gel de collagène tridimensionnel. Quand les deux protéines Hic5 et CRP2 sont coexprimées dans les cellules l'effet inhibiteur sur la contraction est augmenté, et les deux protéines colocalisent sur les fibres de stress d'actine. Sous l'effet de l'étirement cyclique des cellules, les protéines CRP2 et Hic5 formant un complexe sur les fibres de stress, pourraient coopérer dans la régulation de la force contractile des cellules (Kim-Kaneyama *et al.*, 2005).

De la même manière que les protéines CRP1 et CRP2, la protéine CRP3/MLP est colocalisée avec l'actine filamenteuse quand elle est exprimée dans les fibroblastes adhérents (Louis *et al.*, 1997). L'association de la protéine CRP3/MLP cytoplasmique aux filaments d'actine se fait par l'intermédiaire de son domaine LIM C-terminal (Arber & Caroni, 1996). Dans le muscle cardiaque, la protéine CRP3/MLP est concentrée au niveau des stries Z

## Introduction Générale

séparant les sarcomères et pourrait participer avec l' $\alpha$ -actinine à l'ancrage des filaments d'actine sur ces sites (Arber *et al.*, 1997). Dans le cytoplasme, CRP3/MLP est capable de s'associer à travers son domaine LIM C-terminal avec la  $\beta$ I-spectrine et dans les myofibrilles des muscles cardiaques et squelettiques, ces deux protéines colocalisent au niveau des sarcolemmes des stries Z et M (Flick & Konieczny, 2000). Les protéines CRP1 et 2 ne sont pas capables de se lier avec la  $\beta$ I-spectrine, et bien que les trois protéines CRP1-3 fixent des partenaires communs (l' $\alpha$ -actinine et la zyxine), ce résultat montre que les protéines CRP ont des spécificités de partenaires corrélées avec leurs spécificités d'expression tissulaire. La protéine MLP pourrait par son interaction avec la  $\beta$ I-spectrine et l' $\alpha$ -actinine stabiliser et maintenir l'association de l'appareil contractile avec le sarcolemme (Flick & Konieczny, 2000). Par des analyses biophysiques et biochimiques du mutant MLP<sup>-/-</sup>, il a été proposé que la protéine MLP ait un rôle dans la perception de l'étirement mécanique des cellules musculaires cardiaques (Knoll *et al.*, 2002). En effet, les souris MLP<sup>-/-</sup> présentent un défaut majeur restreint à la perception de l'étirement passif du muscle cardiaque probablement lié à un mauvais alignement et une désorganisation des stries Z. Par ailleurs, la protéine MLP/CRP3 est capable d'interagir directement avec la telethonine (T-cap), et colocalise avec celle-ci au niveau des stries Z. T-cap est une protéine des stries Z se fixant à la protéine titine et nécessaire à la fonction du sarcomère. La protéine MLP/CRP3 se fixe à T-cap par son domaine N-terminal court, et la mutation W4R dans MLP, liée à des cardiomyopathies dilatées chez l'Homme, abolit cette interaction et provoque une mauvaise localisation de T-cap. Le complexe MLP/T-cap/titine serait alors un composant clé des stries Z percevant l'étirement mécanique des myocytes cardiaques, et une anomalie de ce complexe conduirait à des cardiomyopathies dilatées et des arrêts cardiaques chez l'homme (Knoll *et al.*, 2002). Enfin, une autre mutation dans la protéine MLP/CRP3 (K69R) située dans la région adjacente au premier domaine LIM et impliquée dans l'interaction avec l' $\alpha$ -actinine est également liée à des cardiomyopathies dilatées chez l'homme (Mohapatra *et al.*, 2003). Cette mutation abolit l'interaction avec l' $\alpha$ -actinine, et induit une altération de la localisation de la protéine MLP/CRP3. Ainsi, au niveau des stries Z des sarcomères, le super complexe zyxine/ $\alpha$ -actinine/ $\beta$ I-spectrine/CRP3/T-cap/titine serait probablement important dans le contrôle des propriétés mécaniques des fibres musculaires.

Ainsi les protéines CRP sont des adaptateurs moléculaires importants permettant la formation de complexes multiprotéiques dans deux compartiments cellulaires différents. Dans le noyau, les protéines CRP régulent l'expression de gènes impliqués dans la différenciation des myoblastes en interagissant avec des facteurs de transcription et dans le cytoplasme, les

protéines CRP assurent le bon fonctionnement des cellules musculaires différenciées en interagissant avec les fibres de stress d'actine et d'autres protéines liées à l'actine telles que la zyxine et la  $\alpha$ -actinine. Les protéines à domaine LIM de plantes sont phylogénétiquement proches des protéines CRP animales et pourraient avoir des fonctions similaires dans le noyau et le cytoplasme des cellules végétales.

#### 4.3. Les protéines à domaine LIM chez les plantes

Le premier gène codant pour une protéine à domaine LIM de plante a été découvert en 1992 chez le tournesol (*Helianthus annuus*) et a été appelé *SF3* (Baltz *et al.*, 1992a ; Baltz *et al.*, 1992b). Le gène *SF3* a la particularité d'être fortement et exclusivement exprimé dans les grains de pollen mature. Depuis, plusieurs autres gènes codant pour des protéines à domaine LIM ont été décrits chez les plantes notamment chez le tabac (*Nicotiana tabacum*), le tournesol et *Arabidopsis thaliana* (Eliasson *et al.*, 2000 ; Kawaoka *et al.*, 2000 ; Sweetman *et al.*, 2000). Les gènes *LIM* de plantes sont soit spécifiquement exprimés dans le pollen, soit exprimés de manière ubiquitaire dans la plante. De plus, comme les protéines CRP, elles ont une double localisation dans la cellule : nucléaire et cytoplasmique.



**Figure 30.** Structure des protéines CRP animales et des protéines à domaine LIM de plantes (Weiskirchen & Günther, 2003).

Représentation schématisée de la topologie des protéines CRP humaines (hCRP1, hCRP2, hCRP3/MLP), et CRP-like de tabac (*Nicotiana tabacum*) (NtWLIM1, NtWLIM2, NtPLIM1a, NtPLIM1b, NtPLIM2) et de tournesol (*Helianthus annuus*) (HaPLIM1, HaPLIM2). Le nom des protéines et le nombre total d'acides aminés sont indiqués à gauche de chaque protéine. Les domaines LIM sont figurés par des boîtes bleues et les positions des régions riches en glycines (GRR) sont représentées par des boîtes de couleur magenta.

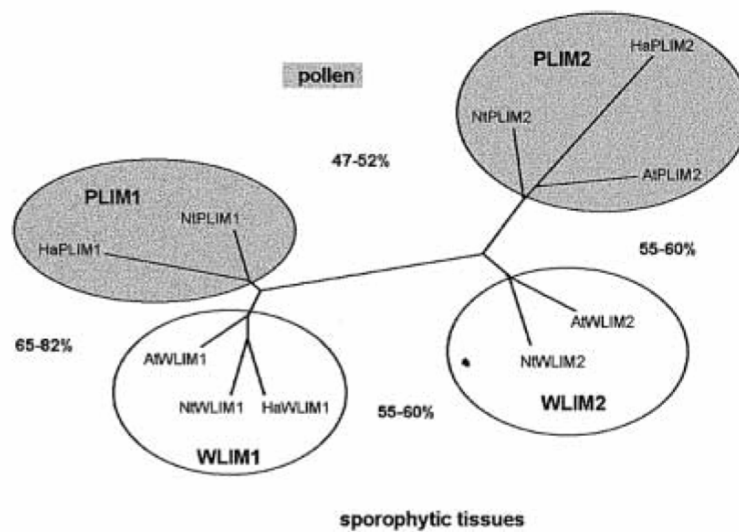
### 4.3.1. Structure des protéines LIM chez les plantes

Les protéines à domaine LIM de plantes ont une taille variant de 181 à 240 acides aminés (Fig. 30) avec une masse moléculaire moyenne de 22 kDa. Ces protéines LIM de plantes sont toutes apparentées et homologues à la famille CRP des protéines LIM animales car elles contiennent deux domaines LIM de 52 résidus très similaires entre eux et ayant la structure caractéristique suivante [C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-H] (Eliasson *et al.*, 2000 ; Mundel *et al.*, 2000). Les protéines à domaines LIM de plantes présentent environ 30 % d'identité en acides aminés avec les protéines CRP1 et 2 humaines, et cette identité de séquence est principalement localisée au sein des deux domaines LIM (Eliasson *et al.*, 2000). Les deux domaines LIM sont séparés par un long espaceur de 40 à 60 acides aminés appelé région inter-LIM. Cette longue région inter-LIM caractérise à la fois les protéines CRP animales et les protéines à domaines LIM de plantes. En effet, les autres protéines LIM animales possèdent en général des domaines LIM très proches entre eux. La région N-terminale est courte avec 10 résidus et la région C-terminale a une longueur variable selon les protéines. Les protéines LIM de plantes diffèrent principalement de leurs homologues CRP animales par l'absence de la région riche en glycine (GRR) suivant chaque domaine LIM chez les protéines CRP, et par leur domaine C-terminal plus long (Fig. 30) (Eliasson *et al.*, 2000 ; Weiskirchen & Günther, 2003). De plus, à l'inverse des protéines CRP contenant une séquence putative d'adressage nucléaire, les protéines à domaine LIM de plantes ne semblent pas contenir de signal de localisation nucléaire (Eliasson *et al.*, 2000). Par ailleurs, les domaines LIM de plantes ont une structure en acides aminés légèrement différente de celle des domaines LIM des protéines CRP animales. En effet, pour chaque domaine LIM de plante, au niveau du dernier ligand au zinc du deuxième doigt de zinc un résidu histidine est trouvé à la place d'une cystéine. De plus, le second domaine LIM est atypique et la structure classique du deuxième doigt de zinc C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-H n'est pas conservée : La cystéine du second ligand du deuxième doigt de zinc est remplacée par une glycine conservée. Il a été suggéré que les deux résidus histidine et cystéine proches de cette glycine servent à la coordination de l'atome de zinc, et deux structures alternatives -C-X-H-X<sub>18</sub>-C-X<sub>2</sub>-H- ou -C-X<sub>4</sub>-C-X<sub>15</sub>-C-X<sub>2</sub>-H- pour la coordination au zinc ont ainsi été proposées (Eliasson *et al.*, 2000 ; Mundel *et al.*, 2000). Enfin, avec l'exception du gène *AtPLIM2* contenant seulement deux introns, les gènes *LIM* de plantes sont généralement interrompus par quatre introns aux positions très conservées. Les premier et quatrième introns sont respectivement localisés dans le premier et le deuxième domaine LIM, et les deuxième et

troisième introns, ayant une courte séquence nucléotidique, sont situés dans la région inter-LIM (Baltz *et al.*, 1992a ; Eliasson *et al.*, 2000 ; Sweetman *et al.*, 2000). Il a été suggéré que les gènes *LIM* des plantes supérieures ont évolué à partir d'une séquence ancestrale ayant la même structure d'épissage d'introns.

4.3.2. Les spécificités structurales des protéines à domaine LIM de plantes sont corrélées avec des profils d'expression distincts

Chez le tabac, *Arabidopsis*, le tournesol et probablement chez toutes les autres plantes supérieures, les protéines à domaine LIM forment deux grands groupes phylogénétiques LIM1 et LIM2 nettement distincts entre eux. Plus précisément, il y aurait quatre isoformes de protéines à domaine LIM correspondant à quatre groupes phylogénétiques différents : PLIM1, PLIM2, WLIM1 et WLIM2 (Eliasson *et al.*, 2000). Les gènes codant pour les protéines PLIM1 et PLIM2 sont fortement et préférentiellement exprimés dans le pollen, et les gènes appartenant aux groupes WLIM1 et WLIM2 seraient plutôt exprimés de manière ubiquitaires dans la plante (Fig. 31).



**Figure 31.** Relations phylogénétiques entre les différentes protéines LIM de plantes (Eliasson *et al.*, 2000)  
L'arbre phylogénétique non enraciné décrit la subdivision de la famille de protéines à domaine LIM de plantes en quatre groupes, deux spécifiquement exprimés dans le pollen (PLIM1 et PLIM2), et deux exprimés dans pratiquement tous les tissus sporophytiques (WLIM1 et WLIM2). Les préfixes d'espèces de plantes sont les suivants : Ha, *Helianthus annuus* ; Nt, *Nicotiana tabacum* ; At, *Arabidopsis thaliana*. Les nombres entre les groupes indiquent l'étendue des identités de séquence (en pourcentage) entre les protéines des différents groupes.



### 4.3.2.1. Les gènes PLIM1 et PLIM2 sont préférentiellement exprimés dans le pollen

Le premier gène *LIM* découvert (*SF3*), plus tard renommé *HaPLIM1a*, a été isolé à partir d'un criblage différentiel d'une banque d'ADNc d'inflorescence de tournesol à l'anthèse, et l'étude de son expression montre que l'ARNm est exclusivement et fortement exprimé dans les grains de pollen, représentant environ 3 % des ARNm totaux (Baltz *et al.*, 1992a ; Eliasson *et al.*, 2000). La protéine SF3-HaPLIM1a possède au niveau de son domaine C-terminal, six répétitions du pentapeptide (S,T,A)(E,D)TQN conférant un caractère acide à ce domaine. Bien qu'aucune étude fonctionnelle n'ait été réalisée, il a été proposé que cette région C-terminale particulière soit un domaine trans-activateur de la transcription (Baltz *et al.*, 1992a). La protéine HaPLIM1a-SF3 est spécifiquement exprimée dans les grains de pollen mature de tournesol (Baltz *et al.*, 1999). D'autres gènes codant pour des protéines PLIM1, orthologues à la protéine HaPLIM1a-SF3, ont été identifiées chez le tabac, le pétunia et *Gerbera hybrida* et tous sont fortement exprimés dans le pollen ou les étamines (Guyon *et al.*, 2000 ; Sweetman *et al.*, 2000 ; Laitinen *et al.*, 2005). Toutefois, aucun gène orthologue aux gènes *PLIM1* n'a encore été identifié chez *Arabidopsis*. L'expression des gènes *NtPLIM1a* et *b* de tabac augmente progressivement et très fortement à partir du stade de pollen bicellulaire jusqu'à un niveau maximum au stade de pollen mature (Sweetman *et al.*, 2000). Chez le pétunia (*Petunia hybrida*), le transcrite *PGPS/D1* homologue au gène *HaPLIM1* est fortement et exclusivement exprimé dans les anthères à un stade tardif de développement et dans le pollen en germination (Guyon *et al.*, 2000). L'expression du gène *PGPS/D1* continue d'augmenter 2 h après la germination. Ainsi, comme pour *HaPLIM1*, il a été suggéré une implication des gènes *NtPLIM1a* et *b* et *PGPS/D1* dans la germination du pollen ou la croissance du tube pollinique. De façon intéressante l'induction de l'expression du gène *PGPS/D1* lors de la germination du pollen est uniquement due à une réponse au kaempférol, un composé de la voie de biosynthèse des flavonoïdes indispensable à la germination du pollen chez le pétunia (Guyon *et al.*, 2000). En effet, même si la germination du pollen est inhibée par l'absence de bore ou de calcium dans le milieu de culture, le transcrite *PGPS/D1* s'accumule fortement dans le pollen après ajout de kaempférol.

D'autres gènes *LIM* spécifiquement exprimés dans le pollen et appartenant au groupe PLIM2 ont été identifiés chez le tabac (*NtPLIM2*), le tournesol (*HaPLIM2*), et *Arabidopsis* (*AtPLIM2*), bien que pour ce dernier l'expression du gène soit détectée dans les feuilles et tiges, et n'ait pas été analysée dans les grains de pollen (Eliasson *et al.*, 2000). Toutefois, des

## Introduction Générale

analyses d'expression de gènes par microarray chez *Arabidopsis* ont révélé une très forte accumulation (d'un facteur 80) de la quantité de transcrite *AtPLIM2* dans les grains de pollen par rapport aux autres tissus (Becker *et al.*, 2003). Par rapport aux autres protéines à domaine LIM de plante, ces trois protéines PLIM2 possèdent un domaine C-terminal relativement long et à caractère acide. De même, chez *Gerbera hybrida*, des transcrits codant pour des protéines à domaines LIM similaires aux protéines PLIM2 (et aussi PLIM1) semblent être exclusivement exprimés dans les étamines (Laitinen *et al.*, 2005). En effet, l'analyse microarray indique que ces transcrits *PLIM1* et *PLIM2* s'accumulent plus de huit fois dans les étamines et ne sont pas du tout exprimés dans les autres tissus. Enfin, une étude microarray a montré qu'un gène codant pour une protéine à domaine LIM de riz est exprimé à un niveau maximum dans le pistil (autopollinisé) à l'anthèse et son expression décroît graduellement jusqu'à 6 h après l'anthèse (Yoshida *et al.*, 2005). Chez le riz, l'élongation du tube pollinique à partir du stigmate jusqu'au micropyle dure 1 heure, puis la fertilisation du micropyle se déroule entre 2 et 4 heures après l'anthèse. L'expression de ce gène *LIM* dans d'autres tissus suggère toutefois des fonctions additionnelles, non liées au processus d'élongation du tube pollinique. Bien que les protéines PLIM1 et PLIM2 soient probablement impliquées dans des interactions avec l'actine, leur fonction précise lors de la germination du pollen ou la croissance du tube pollinique n'a pas encore été déterminée.

### 4.3.2.2. Les gènes WLIM1 et WLIM2 sont exprimés de manière ubiquitaire

Trois gènes *WLIM1* ont été identifiés chez *Arabidopsis* (*AtWLIM1*), le tabac (*NtWLIM1*) et le tournesol (*HaWLIM1*), et ces gènes apparaissent exprimés dans tous les tissus de la plante (Eliasson *et al.*, 2000 ; Kawaoka *et al.*, 2000 ; Mundel *et al.*, 2000). La protéine EcLIM1 d'eucalyptus (*Eucalyptus camaldulensis*) présente 84 % d'homologie avec la protéine NtLIM1 (*NtWLIM1*) de tabac, et l'expression des gènes *Eclim1* et *NtLIM1* est plus forte dans les tiges que dans les feuilles (Kawaoka *et al.*, 2000 ; Kawaoka *et al.*, 2006). La protéine HaWLIM1 de tournesol présente respectivement 70 % et 80 % d'identité de séquence avec la protéine HaPLIM1-SF3 et son orthologue ATL1 (*AtWLIM1*) d'*A. thaliana* (Mundel *et al.*, 2000). Contrairement à HaPLIM1-SF3, les protéines HaWLIM1 et ATL1 ne possèdent pas la région C-terminale composée de six répétitions du pentapeptide (S,T,A) (E,D) TQN. Le niveau d'expression du gène *HaWLIM1* est plutôt faible. Toutefois, les transcrits *HaWLIM1* et la protéine HaWLIM1 sont détectés dans une grande variété d'organes

## Introduction Générale

comme l'hypocotyle, les tiges, les feuilles, et les organes reproducteurs mâle et femelle incluant le pollen et les ovaires (Mundel *et al.*, 2000). Au niveau tissulaire, les analyses d'immunohistolocalisation montrent que la protéine est fortement exprimée dans les racines au niveau de la région méristématique apicale, et des cellules du protoderme. La protéine est également détectée dans les tissus vasculaires du cylindre central. Dans la tige, HaWLIM1 est localisée dans les cellules épidermiques, dans les cellules plus internes du cortex, dans le cambium interfasciculaire et dans les tissus vasculaires du phloème (Mundel *et al.*, 2000). Chez *Arabidopsis*, le gène *AtWLIM1* semble exprimé dans toute la plante et les transcrits sont plus abondants dans les tiges que dans les feuilles et fleurs (Eliasson *et al.*, 2000). De même, les transcrits *BnLIM* de colza (*Brassica napus*) sont trouvés dans tous les organes végétatifs et reproducteurs (Huang *et al.*, 2007). De plus, l'expression du gène *BnLIM* augmente en fonction du remplissage de la graine avec un maximum d'expression à 23-29 jours après l'anthèse et est induite dans la plantule après l'ajout de saccharose. La protéine BnLIM pourrait ainsi être impliquée dans le développement de la graine (Huang *et al.*, 2007).

Les gènes *AtWLIM2* et *NtWLIM2* ont été isolés chez *Arabidopsis* et le tabac, et semblent comme les gènes *WLIM1* exprimés de manière ubiquitaire dans la plante (Eliasson *et al.*, 2000). Toutefois, chez le tournesol le gène orthologue *HaWLIM2* n'a pas été identifié. Chez *Arabidopsis*, le gène *AtL2* codant pour une protéine à domaine LIM appartenant probablement au sous-groupe WLIM2 est plus exprimé dans les racines que dans les tiges (Tranbarger *et al.*, 2003). De plus, l'expression du gène semble diminuer dans les racines quand la quantité en azote du milieu de culture augmente. De façon surprenante, le gène *AtWLIM2* est exprimé de manière préférentielle dans les siliques de la même manière que le gène *FUS3*, un gène *LEAFY COTYLEDON (LEC)* codant pour un régulateur de transcription favorisant l'embryogenèse (Harada, 2001 ; Dean Rider *et al.*, 2003). De plus, l'expression du gène *AtWLIM2* est induite dans les graines en germination chez le mutant *pkl* (Dean Rider *et al.*, 2003). La protéine PICKLE (PKL), code pour un facteur putatif CHD3 de remodelage de la chromatine et agit comme un répresseur de l'identité embryonnaire en régulant négativement la transcription des gènes durant la germination des graines. Il a ainsi été suggéré que le gène *AtWLIM2* soit un régulateur du développement embryonnaire d'une manière analogue aux gènes *LEC*, des régulateurs de transcription.

D'autres analyses d'expression chez les plantes ont identifié d'autres gènes *LIM* ayant des profils d'expression particuliers, mais l'appartenance de ces gènes à un sous-groupe phylogénétique particulier n'a pas encore été déterminée. Par exemple, une étude d'expression chez *Arabidopsis* a identifié un gène codant pour une protéine à domaine LIM

dont l'expression augmente progressivement au cours du processus de sénescence des feuilles (Gepstein *et al.*, 2003). Par ailleurs, de manière intéressante en regard de la formation du bois chez le peuplier, un autre transcrite codant pour une protéine à domaine LIM chez le pin (*Pinus taeda*) est plus fortement exprimé dans le xylème que dans d'autres tissus de l'arbre tels que le phloème et les feuilles (Allona *et al.*, 1998). De plus, à l'inverse des observations obtenues chez le peuplier montrant une probable induction de l'expression de gènes codant pour des protéines LIM dans le bois de tension, l'accumulation du transcrite *LIM* chez le pin semble plus forte dans le xylème ne formant pas de bois de compression. Enfin, une étude d'expression récente chez le peuplier montre que deux transcrits *LIM* respectivement similaires aux gènes *WLIM1* et *PGPS/D1* sont plus fortement exprimés dans les tiges à croissance secondaire et le xylème secondaire que dans l'écorce, les feuilles et les tiges à croissance primaire (Prassinis *et al.*, 2005).

### 4.3.3. Les protéines LIM subissent des modifications post-traductionnelle

De façon surprenante, les analyses des protéines LIM de plante en Western-blot montrent souvent la présence de deux polypeptides migrant à des tailles inattendues, supérieures à celles prédites. En effet, deux polypeptides HaPLIM1a sont détectés dans les extraits protéiques de pollen vers 30 et 50 kDa suggérant la formation d'un dimère de protéines LIM ou des modifications post-traductionnelles (Baltz *et al.*, 1999). De même, alors que la taille attendue de la protéine HaWLIM1 est de 21 kDa, celle-ci migre en western-blot vers une taille de 50 kDa suggérant également des modifications post-traductionnelles ou une dimérisation de la protéine. Un polypeptide additionnel de 34 kDa est quelquefois détecté, et le rapport entre ces deux polypeptides de tailles différentes varie probablement en fonction du stade de développement de l'organe analysé. En effet, les ovaires jeunes contiennent majoritairement le polypeptide de 34 kDa, alors que dans les ovaires matures c'est le polypeptide de 50 kDa qui prédomine (Mundel *et al.*, 2000). Ces anticorps anti-GST-HaWLIM1 utilisés en Western blot sur des extraits de protéines totales issus de protoplastes de tournesol ou de bactéries exprimant la protéine recombinante HaWLIM1 détectent principalement deux peptides vers 27 kDa et 52 kDa (Brière *et al.*, 2003). Par contre, sur ces mêmes extraits protéiques, lorsque des anticorps dirigés contre des peptides de synthèse P16 et P14 respectivement localisés dans les régions inter-LIM et C-terminale de la protéine HaWLIM1 sont utilisés, seul le polypeptide migrant vers 27 kDa est détecté. Ces anticorps ne

reconnaissent probablement pas les homo ou hétérodimères de protéines LIM. Toutefois la taille des protéines LIM déterminée expérimentalement est supérieure à celle prédite suggérant des modifications post-traductionnelles.

#### *4.4. Les protéines LIM de plante ont une double localisation subcellulaire*

Comme pour les protéines CRP animales, les protéines LIM de plantes sont à la fois localisées dans le noyau et le cytoplasme. La colocalisation et l'accumulation simultanée de la protéine HaPLIM1 avec les filaments d'actine dans la région corticale des cônes germinatifs des grains de pollen (pollen tricellulaire), suggèrent la participation de SF3 dans la croissance du tube pollinique (Baltz *et al.*, 1999). En effet, le remodelage du cytosquelette d'actine et la formation de câbles d'actine permettant le flux cytoplasmique est important pour l'élongation du tube pollinique à son extrémité (Vidali & Hepler, 2001). Par ailleurs, la protéine SF3 pourrait, comme les protéines CRP animales, participer à des interactions entre le cytosquelette d'actine et la membrane plasmique (Baltz *et al.*, 1999). D'autre part, les analyses d'immunohistocalisation montrent que la protéine HaWLIM1 est localisée à la fois dans le cytoplasme et/ou dans le noyau selon le type cellulaire (Mundel *et al.*, 2000 ; Brière *et al.*, 2003). Dans le cytoplasme, la protéine HaWLIM1 est souvent associée avec les plastes et d'autres organites intracellulaires plus petits ressemblant à des vésicules membranaires. De façon intéressante, dans le cortex des tiges, la protéine HaWLIM1 est associée aux amyloplastes, des organites cellulaires encore appelés statolithes et capables de sédimenter sous l'effet de la gravité. Les amyloplastes attachés par des filaments d'actine sont importants dans la perception du signal gravitationnel des tiges et des racines (Blancaflor & Masson, 2003), et l'association de la protéine HaWLIM1 à ces structures indique probablement un rôle dans la perception du gravitropisme de la plante. La protéine HaWLIM1 peut aussi former une couche dense proche des plaques criblées des éléments du phloème (Mundel *et al.*, 2000). Il est intéressant de noter que chez le peuplier, l'actine et la myosine sont abondantes sur le pourtour des plaques criblées du phloème (Chaffey & Barlow, 2002). Dans les cellules en division au stade anaphase/télophase, la protéine HaWLIM1 est aussi associée au phragmoplaste (une structure composée de tubuline et d'actine), indiquant une possible implication dans la formation de la plaque cellulaire (Mundel *et al.*, 2000). L'association fréquente de la protéine HaWLIM1 avec des organites intracellulaires et sa localisation au niveau du phragmoplaste suggèrent également une fonction dans l'ancrage du cytosquelette

d'actine à la membrane plasmique et une implication dans le flux cytoplasmique. Contrairement aux observations précédentes, l'analyse en microscopie confocale sur des protoplastes de tournesol montre qu'en fin de télophase la protéine HaWLIM1 n'est que faiblement associée au phragmoplaste (Brière *et al.*, 2003). Enfin, la protéine NtWLIM1 fusionnée avec la GFP et exprimée dans les cellules BY2 de tabac a une localisation nucléaire et cytoplasmique (Kawaoka & Ebinuma, 2001 ; Thomas *et al.*, 2006). De même, la protéine BnLIM-GFP est détectée dans le cytoplasme et le noyau des cellules épidermiques d'oignon (Huang *et al.*, 2007). Cette double localisation subcellulaire des protéines LIM de plante suggère une double fonction cellulaire dans le noyau comme une protéine du complexe de transcription, et dans le cytoplasme comme une protéine interagissant avec le cytosquelette d'actine.

#### 4.4.1. Dans le noyau, la protéine NtWLIM1 agit comme un facteur de transcription impliqué dans la voie de biosynthèse des phénylpropanoïdes

Chez le tabac, NtLIM1 (NtWLIM1) a été découvert lors d'un criblage de protéines se fixant sur la boîte *PAL* (*PAL-box*), en utilisant une banque d'expression d'ADNc isolés à partir de tiges (Kawaoka *et al.*, 2000). Comme nous l'avons vu précédemment, la boîte *PAL* est un motif riche en AC du promoteur *prxC2* ressemblant à un élément *cis* activateur de type ACII impliqué dans l'expression de gènes appartenant à la voie de biosynthèse des phénylpropanoïdes. Cette boîte *PAL* est essentielle pour le niveau basal de l'expression du gène *prxC2* codant pour une peroxydase de raifort, ainsi que pour l'induction de l'expression de ce gène en réponse à la blessure (Kaothien *et al.*, 2002). De plus, la protéine NtLIM1 entière et chacun des deux domaines LIM individuels sont capables de se lier à l'ADN (boîte *PAL*) *in vitro* et *in vivo* (Kawaoka *et al.*, 2000). La séquence consensus de fixation de la protéine NtLIM1 sur l'ADN a été déterminée et est la suivante : [CCAC(A/C)AN(A/C)N(C/T)(A/C)] où N représente n'importe quel nucléotide. Ainsi, les domaines LIM auraient probablement la même fonction que les domaines à doigt de zinc de type GATA-1 qui jouent à la fois un rôle de liaison à l'ADN et d'interaction protéine/protéine. Cependant, alors que la protéine entière est capable d'activer l'expression du gène rapporteur *GUS* fusionné à la boîte *PAL* dans des protoplastes de tabac, les domaines LIM seuls ne sont pas suffisants pour activer la transcription. Ainsi, la protéine NtLIM1 agit certainement comme un facteur de transcription activateur *in vivo*, et il est probable que le domaine acide en C-terminal de la protéine joue un rôle d'activateur de transcription. D'autre

part, la protéine NtLIM1 est nécessaire à l'expression basale et inductible par la blessure du promoteur *prxC2* (Kaothien *et al.*, 2002). Toutefois, la protéine NtLIM1 se fixe à la boîte *PAL* indépendamment du stress blessure, et l'expression du gène *NtLIM1* n'est pas induite dans les feuilles blessées (Kawaoka *et al.*, 2000 ; Kaothien *et al.*, 2002). Ainsi, NtLIM1 n'est probablement pas un facteur de transcription répondant à la blessure. Par ailleurs, des lignées de tabac transgéniques portant la construction antisens *NtLIM1* montrent une diminution de l'expression des gènes *NtLIM1*, *PAL*, *4CL* et *CAD*. De plus, l'activité enzymatique des protéines *PAL*, *4CL* et *CAD* est aussi diminuée, et une réduction jusqu'à 27% de la teneur en lignine est observée chez la lignée la plus affectée (Kawaoka *et al.*, 2000). Le rapport syringaldéhyde/vanilline diminue chez les plantes antisens *Ntlim1*, indiquant une modification de la composition en lignine avec probablement une augmentation d'unités guaiacyles et une diminution d'unités syringyles. De même, une construction antisens *NtLIM1* introduite chez l'eucalyptus visant à supprimer l'expression du gène orthologue *Eclim1*, provoque une diminution de l'expression des gènes *PAL*, *C4H* et *4CL*, et une réduction de l'activité *4CL* vis à vis des acides féruliques (Kawaoka *et al.*, 2006). De plus, une réduction de 29 % de la teneur en lignine et une augmentation de cellulose et d'hémicellulose sont observées chez la lignée transgénique la plus affectée dans l'expression du gène *Eclim1*. Ces résultats suggèrent un rôle des protéines NtLIM1 et EclIM1 dans l'activation de la transcription des gènes *PAL*, *C4H*, *4CL* et *CAD* et dans la régulation de la voie de biosynthèse des lignines.

Concernant les autres protéines LIM de plante, leur rôle dans la transcription des gènes n'a pas encore été démontré. Toutefois, la protéine HaPLIM1a-SF3 serait capable de se fixer *in vitro* aux acides nucléiques (DNA et/ou RNA) de manière non spécifique (Baltz *et al.*, 1996). En effet, la protéine *HaPLIM1a-SF3* est capable de se fixer sur son propre promoteur (nucléotides -240 à +40) conférant une expression spécifique du pollen, et sur des fragments de 20 pb de ce promoteur indiquant une fixation de la protéine sur des séquences d'ADN aléatoires. L'utilisation de polypeptides HaPLIM1a-SF3 mutés indique que les régions inter-LIM et C-terminale ne participent pas à la fixation avec l'ADN, alors que les deux domaines LIM sont tous les deux nécessaires à cette interaction. La protéine HaPLIM1a-SF3 est également capable de lier différents types d'ARN (Baltz *et al.*, 1996). D'autre part, la protéine AtWLIM1 serait capable comme HaPLIM1/SF3 de se lier aux acides nucléiques de manière non spécifique (Mundel *et al.*, 2000). Enfin, l'ADNc *NtWLIM2* a été isolé après criblage d'une banque d'expression avec une sonde ADN contenant le promoteur du gène *H1*, et une capacité de fixation à l'ADN n'est donc pas exclue pour la protéine NtWLIM2 (Mundel *et al.*, 2000).

### 4.4.2. Dans le cytoplasme, la protéine NtWLIM1 se lie et empaquette les filaments d'actine

Bien qu'il n'ait jamais été encore montré que les protéines CRP animales soient capables de s'associer au cytosquelette de microtubules, la protéine HaWLIM1 de tournesol pourrait être associée indirectement aux filaments de microtubules (Brière *et al.*, 2003). Des études d'immunolocalisation ont montrées que dans des protoplastes de tournesol en interphase, la protéine HaWLIM1 est distribuée sous forme de structures ponctuées le long des microtubules corticaux. En effet, il est observé une forte colocalisation de la protéine HaWLIM1 avec la  $\beta$ -tubuline. De plus, le traitement des protoplastes avec l'oryzaline une drogue dépolymérisant les microtubules affecte la localisation de la protéine HaWLIM1, alors que la latrunculine B un inhibiteur de la polymérisation de l'actine n'a pas d'effet sur la distribution cellulaire de la protéine. Actuellement, aucune propriété de fixation aux microtubules n'a été démontrée pour les protéines CRP animales, alors qu'il y a de forts arguments supportant une liaison de ces protéines à l'actine (Grubinger & Gimona, 2004 ; Tran *et al.*, 2005). Ces résultats montrant une liaison de la protéine HaWLIM1 au moins indirecte avec les microtubules sont donc surprenants. Toutefois, deux types de protéines animales contenant des domaines LIM, la paxilline et la LIM kinase 1 (LIMK1), interagissent directement avec la tubuline (Herreros *et al.*, 2000 ; Gorovoy *et al.*, 2005). Mais, alors que la paxilline interagit avec les microtubules par l'intermédiaire de ces quatre domaines LIM, pour la LIMK1 cette interaction est liée au domaine PDZ de la protéine.

Plus récemment, une étude a montré que la protéine NtWLIM1 de tabac est une nouvelle protéine se fixant à l'actine (ABP) comme les protéines CRP1 et CRP2 (Thomas *et al.*, 2006). Dans les cellules BY2 de tabac, la protéine NtWLIM1 fusionnée à la GFP est détectée à la fois dans le noyau et le cytoplasme où elle est colocalisée avec le cytosquelette d'actine. Pendant l'interphase, la protéine NtWLIM1-GFP décore les filaments d'actine et lors de la télophase, elle est comme l'actine associée au phragmoplaste. Contrairement, aux observations précédentes montrant une association de la protéine HaWLIM1 de tournesol avec le réseau de microtubules, la protéine NtWLIM1-GFP ne colocalise pas avec les microtubules durant le cycle cellulaire, et le traitement avec l'oryzaline n'affecte pas sa localisation dans la majorité des cellules. Par contre, lorsque les cellules sont traitées avec la latrunculine B, à la fois l'organisation cellulaire de l'actine et de la protéine NtWLIM1 est totalement perturbée. Cette différence de localisation des protéines WLIM1 de tournesol et de



tabac serait imputable à l'utilisation de détergents induisant des artéfacts de localisation de la protéine avec les microtubules (Thomas *et al.*, 2006). Confortant cette hypothèse, des expériences d'assimilation *in vivo* d'anticorps anti-HaWLIM1 dans les protoplastes de tournesol (sans détergent) montrent que ces anticorps induisent une désorganisation du réseau d'actine dans les cellules (Brière *et al.*, 2004). Les expériences de cosédimentation à haute vitesse montrent que la protéine recombinante NtWLIM1-6His s'associe directement avec l'actine filamenteuse avec une forte affinité et la stoechiométrie de cette interaction à saturation est de 1,9 mole de protéine WLIM1 liée à 1 mole d'actine (Thomas *et al.*, 2006). De plus, il a été montré que la protéine NtWLIM1 stabilise les filaments d'actine contre la dépolymérisation par la latrunculine B *in vivo* et *in vitro*. En effet, la dépolymérisation de l'actine par la latrunculine B est retardée dans les cellules BY2 transformées avec la protéine NtWLIM1-GFP. De plus, les mêmes résultats ont été obtenus par des expériences de cosédimentation à haute vitesse, où l'ajout de protéines NtWLIM1 protège l'actine filamenteuse de l'effet de la latrunculine B. Par ailleurs, lorsque la protéine NtWLIM1-GFP est exprimée dans les cellules épidermiques de feuille de tabac, les filaments et câbles d'actines sont moins nombreux et plus épais. De plus, des expériences de cosédimentation à basse vitesse et des observations microscopiques indiquent que la protéine WLIM1 seule est capable *in vitro* d'empaqueter les filaments d'actine en des câbles épais (Thomas *et al.*, 2006). Ces résultats montrent que la protéine NtWLIM1 se fixe indépendamment aux filaments d'actine sans l'intermédiaire d'autres protéines et augmente la stabilité des filaments d'actine en favorisant l'empaquetage de ces filaments en câbles d'actine. Ainsi, avec les villines et fimbrines (Kovar *et al.*, 2000 ; Huang *et al.*, 2005), la protéine WLIM1 forme une troisième famille de protéine empaquetant l'actine chez les plantes. Cette capacité de la protéine WLIM1 à former des câbles de filament d'actine épais est intéressante en regard des protéines PLIM1 ou PLIM2 qui pourraient ainsi participer à la formation et le maintien du tube pollinique en stabilisant les câbles d'actine. Enfin l'expression hétérologue de la protéine NtWLIM1 fusionnée à la GFP dans les cellules Vero de singe montre une colocalisation nette avec l'actine, suggérant une possible similarité de fonction entre les protéines LIM de plante et CRP animales (Thomas *et al.*, 2007). Toutefois, des résultats préliminaires indiquent que la protéine NtWLIM2 de tabac n'est pas associée de manière évidente avec l'actine dans les cellules BY2 et épidermique de tabac. Les différentes protéines LIM peuvent donc avoir diverses fonctions cytoplasmique et nucléaire dans la plante.

Certains gènes *LIM* de peuplier semblent fortement exprimés dans le xylème secondaire et plus particulièrement dans le bois de tension (Déjardin *et al.*, 2004 ; Prassinis *et*

## Introduction Générale

*al.*, 2005 ; Andersson-Gunneras *et al.*, 2006). Or le nombre de gènes *LIM* chez les espèces du genre *Populus* est actuellement inconnu. De plus, l'organisation phylogénétique de la famille de protéines à domaine LIM de peuplier, ainsi que les spécificités d'expression des gènes *LIM* dans les différents tissus de l'arbres n'ont pas encore été déterminés. Certains gènes *LIM* de peuplier pourraient former de nouveaux sous-groupes phylogénétiques ayant des spécificités d'expression dans les tissus vasculaires du xylème. Par ailleurs, la phylogénie de cette famille de protéines à domaine LIM chez les plantes est incomplète. En effet, il est fort probable que le génome d'*Arabidopsis* contienne d'autres gènes *LIM* en plus de ceux déjà identifiés (*AtWLIM1*, *AtWLIM2* et *AtPLIM2*), et les gènes *LIM* de riz, une monocotylédone, n'ont toujours pas été décrits.

## 5 Objectifs de la thèse et démarche expérimentale

Le peuplier est devenu l'espèce modèle pour la génomique fonctionnelle des arbres et l'étude de la formation du bois. En effet, plus de 350000 séquences ESTs de peuplier sont actuellement disponibles dans les bases de données publiques et le séquençage du génome du peuplier est maintenant achevé. Nous avons vu précédemment que beaucoup de gènes impliqués dans divers processus physiologiques sont dérégulés dans le bois de tension. Au centre INRA d'Orléans, 10062 ESTs ont été séquencées à partir d'échantillons prélevés sur du bois tendu et opposé à différents stades de différenciation (Déjardin *et al.*, 2004). Certains transcrits codant pour des protéines à domaine LIM sont plus abondants dans le xylème de bois de tension par rapport au bois opposé ou normal (Déjardin *et al.*, 2004 ; Andersson-Gunneras *et al.*, 2006). Ces protéines sont donc des candidats intéressants potentiellement impliqués dans la régulation de la formation de ce bois particulier.

**L'objectif de ma thèse est d'analyser et de comprendre la contribution de chacun des membres de la famille de protéines à domaine LIM dans les différents aspects du développement et de la physiologie du peuplier, notamment lors de la formation du bois de tension. Ces données permettront d'appréhender puis de caractériser la fonction de chacun des gènes *LIM* lors du processus complexe de la différenciation du bois.**

Tout d'abord, pour obtenir un inventaire des gènes codant pour des protéines à domaine LIM chez le peuplier, je poursuivrai la recherche de séquences nucléotidiques *LIM* dans les différentes banques d'ADNc de peuplier et le génome de *Populus trichocarpa*. Afin de mieux comprendre l'organisation de la famille de protéines à domaine LIM chez le peuplier et les plantes, cette recherche sera étendue à d'autres espèces, notamment *Arabidopsis* et le riz dont les génomes sont séquencés. La phylogénie de cette famille chez les plantes sera déterminée et des motifs conservés ayant une fonction putative dans la régulation de ces protéines seront recherchés. La distribution des ESTs *LIM* entre les différentes banques d'ADNc sera ensuite analysée. Pour comprendre et avoir une idée plus précise de la régulation de l'expression des gènes *LIM* de peuplier, des études d'expression par RT-PCR semi quantitative seront conduites sur différents tissus de l'arbre, notamment le pollen et le bois de tension. Enfin, pour le gène le plus intéressant, la production et la purification de protéines recombinantes LIM sera envisagée afin d'obtenir des anticorps spécifiques. Ces anticorps nous seront utiles pour diverses applications comme la détection de la protéine dans différents tissus par Western blot et dans les cellules du xylème par immunocyto-localisation.

## Partie 1. Analyse génomique et phylogénétique de la famille de protéines à domaine LIM chez le peuplier et les plantes

### 1 Matériel et méthodes

#### 1.1. Caractérisation des séquences d'ADNc et de gènes LIM de peuplier

Les transcrits codant pour des protéines similaires aux protéines à domaines LIM (contenant uniquement deux domaines LIM) ont été recherchés au sein d'une collection de 10062 ESTs obtenues à partir de banques d'ADNc de bois de *Populus tremula* × *P. alba* (clone INRA #717-1-B4, section *Populus*) (Déjardin *et al.*, 2004). Les ESTs ont été regroupées par similarité de séquences en cluster, contigs, puis consensus grâce au logiciel StackPACK (Miller *et al.*, 1999 ; Déjardin *et al.*, 2004). Cet assemblage des ESTs du projet Lignome est hébergé sur le site du Centre Bioinformatique de Bordeaux (<http://cbi.labri.fr/outils/SAM2/>) Dans un premier temps, une recherche a été effectuée sur la base de l'annotation des séquences dans la base de données d'ESTs « SAM2 » (login : guest, password : guest, projet : « peuplier2 ») en utilisant comme requête les mots clés « SF3 », « LIM transcription factor » et « LIM domain ». Puis le programme « Basic Local Alignment Search Tool » (BLAST) (Altschul *et al.*, 1990) a été utilisé pour rechercher, dans la base de données d'ESTs « ForEST » développée dans notre unité et celle d'ESTs de plante (*Viridiplantae*) du NCBI, d'autres séquences d'ESTs similaires aux séquences LIM déjà identifiées par mot clé. Le programme BLAST permet par des alignements de séquences de trouver des régions de similitude locale entre les séquences nucléotidiques ou protéiques. Parmi les ESTs *PtaLIM* groupées dans un même consensus et pouvant correspondre à un même gène, nous avons sélectionné le clone ADNc le plus long pour le séquencer. (Les tailles des inserts d'ADNc avaient été estimées après analyse sur gel d'agarose.) Les ADNc entiers, portés par le vecteur pTriplex (Clontech, Laboratories Inc., Mountain View, CA, USA), ont été séquencés en utilisant l'amorce sens TriplexA (5'-CTCGGGAAGCGCGCCATTGTGTTGG-3') et l'amorce antisens TriplexB1 (5'-ATACGACTCACTATAGGGCGAATTGGC-3'). Nous avons complété le séquençage des ADNc en 3' et 5' avec de nouvelles amorces situées dans l'ADNc et spécifiques pour chaque gène *PtaLIM* (Tableau 7).

## Partie 1. Analyse phylogénétique

**Tableau 7** Amorces oligonucléotidiques utilisées soit pour compléter les séquences des ESTs *PtaLIM* afin d'obtenir une séquence d'ADNc pleine longueur, soit pour séquencer les produits RT-PCR afin de corriger et améliorer l'annotation des séquences génomique de *Populus trichocarpa*.

Gène	Clone d'ADNc ou produits RT-PCR	Amorces sens et anti-sens (5' to 3')
<i>PtaWLIM1a</i>	CF231651	GACAAGTGTTCGGCTGCAAGAAC et AGGCAGGAAAATGCAGGATGAACA
<i>PtaWLIM1b</i>	CF236568	GACAAGTGTTCGGCTGCAAGAAC et GCAAGGCAGGAAAATACACTCGTT
<i>PtaGLIM1a</i>	CF235176	CCTGTTGATAATGAGAACGCCAGC et CAGCAATCTCTACAGCAGTGATGC
<i>PtaGLIM1b</i>	CF236473	CCTGTTGATAATGAGAACGCAAGC et CACGGCCGGTGATATTCACAGTG
<i>PtaWLIM2a</i>	CF235467	GGTGATCCGAGACAGGTAACAGT et GAGATGTTTAAGCCTCTGGCACAG
<i>PtaPLIM2a</i>	CF236773	CATTCACGGGAACCCTGGATAAG et TGCTCTTCTGACTTTTCCTCCTCC
<i>PtaPLIM2b</i>	EF151463	CATTCACGGGAACCCTGGATAAG et TCGGCCAGTTCAGGAGGTGG
<i>Pta<math>\delta</math>LIM2a</i>	EF151464	GGACCTTAGAGAAATGCAAGG et GATACCATCAAGAGCTGCATAA
<i>Pta<math>\delta</math>LIM2b</i>	EF151465	GGACCTTAGAGAAATGCAAGG et CAAGAGCAGCGTAGGATGAT

Les différents gènes *PtLIM* chez *Populus trichocarpa* (section *Tacamahaca*) ont été identifiés dans la base de données du génome de peuplier ([http://genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html)) (Tuskan *et al.*, 2006) soit par recherche de mots clés, en utilisant comme requête l'annotation InterPro du domaine LIM (IPR001781), soit en utilisant les six ADNc *PtaLIM* isolés à partir de *Populus tremula*  $\times$  *P. alba* pour des recherches de séquences par BLAST (BlastN). Enfin, les gènes encodant des protéines à deux domaines LIM ont été sélectionnés. Pour chaque gène, la prédiction des jonctions exon/intron a été vérifiée en comparant la séquence génomique prédite avec la séquence des ADNc obtenus expérimentalement et celle des ESTs et ADNc disponibles dans la base de données du NCBI. Quand les ADNc n'étaient pas disponibles (gènes *PtPLIM2b*, *Pt $\delta$ LIM2a* et *Pt $\delta$ LIM2b*), les positions prédites d'épissage des introns ont été vérifiées grâce au séquençage de produits de RT-PCR avec des amorces spécifiques de chaque gène *LIM* de peuplier (Tableau 7). Ces fragments ont été produits après transcription inverse d'ARN polyA préparés à partir d'échantillons de bois (Cf. Matériels et Méthodes Partie 2) et amplification par PCR. Les produits de RT-PCR ont été clonés dans le vecteur pGEM<sup>®</sup>-T easy (Promega) en suivant les instructions du fabricant, et les séquences de région codante partielle des ADNc *PtPLIM2b*

## Partie 1. Analyse phylogénétique

(EF151463), *Pt $\delta$ LIM2a* (EF151464) et *Pt $\delta$ LIM2b* (EF151465) ont été soumises dans la base de données Genbank (<http://www.ncbi.nlm.nih.gov/>).

Finalement, pour compléter la liste des ESTs *LIM* de peuplier disponibles, la base de données d'ESTs PopulusDB (<http://www.populus.db.umu.se>) (Sterky *et al.*, 2004) a été interrogée dans un premier temps avec les mots clés « LIM domain », « LIM transcription factor » et « SF3 ». Les banques d'ADNc de la base de données PopulusDB ont été construites à partir d'échantillons issus d'espèces de peuplier *Populus tremula* ou *P. tremula*  $\times$  *tremuloïdes* (section *Populus*) (Sterky *et al.*, 2004). Les ESTs et ADNc *LIM* de peuplier ont également été recherchés par BlastN dans les bases de données d'ESTs et non redondante (nr) de plantes (*Viridiplantae*) du NCBI et celle de PopulusDB. Cette recherche BLAST a été réalisée jusqu'à Août 2006, en utilisant les paramètres par défaut et comme requête pour chaque gène *LIM*, les séquences de transcrits déduites du génome de *P. trichocarpa*. Ainsi, grâce aux informations de séquence des transcrits pleine longueur ou partiels obtenus au laboratoire et par recherche BLAST, les séquences génomiques prédites des gènes *PtLIM* ont été corrigées par alignement de séquence en utilisant le logiciel ClustalW (Chenna *et al.*, 2003) sur le site de l'EMBL-EBI (<http://www.ebi.ac.uk/clustalw/>). Les séquences génomiques corrigées et actualisées sont disponibles en annexe 1.

### 1.2. Recherche de séquences codant pour des protéines à deux domaines LIM chez les plantes

Jusqu'à Août 2006, j'ai utilisé différents moyens pour rechercher dans différentes bases de données toutes les protéines LIM de plantes contenant uniquement deux domaines LIM. Premièrement, les gènes et ADNc pleine longueur connus et encodant des protéines à domaine LIM chez *Arabidopsis*, le tournesol et le tabac, ont été collectés à partir de la littérature (Eliasson *et al.*, 2000), et la base de données du NCBI. Les gènes et ADNc pleine longueur additionnels annotés comme « LIM domain protein » ou « LIM transcription factor » ont été trouvés par recherche de mot clé, ou en utilisant comme requête l'annotation InterPro du domaine LIM (IPR001781). Alternativement, les séquences des protéines LIM de peuplier ont été utilisées pour une recherche par BLAST de séquences nucléotidiques traduites en protéines (TBlastN) et de séquences protéiques (BlastP) dans la base de données du NCBI non redondante. Chez *Arabidopsis thaliana* et le riz (*Oryza sativa*), les séquences génomiques des gènes *LIM* et celles des ADNc correspondants ont été récupérées dans les bases de données du NCBI, TIGR plant genome (<http://plantgenomics.tigr.org/>), et TAIR

(<http://arabidopsis.org>). La séquence de l'ADNc *EgWLIM2* (clone CPHO63) (Jaligot *et al.*, 2002), m'a été fournie par Tranbarger T. et Beulé T. (IRD, Montpellier). Finalement pour compléter les séquences codant pour des protéines à deux domaines LIM, les ESTs homologues de plantes ont été recherchés par BlastN dans la base de données d'ESTs du NCBI avec comme requête les transcrits de peuplier pleine longueur isolés au laboratoire ou ceux déduits de la séquence génomique. Le processus de recherche BlastN a été répété avec chaque nouvelle séquence de transcrit *LIM* trouvée jusqu'à ce qu'aucune autre séquence présentant une homologie significative (Score  $\geq 50$  bits ; E value  $\leq 1e^{-10}$  ; Max ident  $\geq 50$  % ; Query coverage  $\geq 40$  %) ne soit identifiée. Pour chaque gène, l'EST la plus longue a été traduite en protéine dans tous les cadres de lecture en utilisant le programme EMBOSS Transeq sur le site de l'European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI) (<http://www.ebi.ac.uk/emboss/transeq/>). Afin de s'assurer que ces ESTs codent bien pour une protéine entière, seules les ESTs portant une séquence codante (CDS) complète du codon d'initiation ATG aux codons stops, et des portions de séquences en 3' et 5' de régions non traduites (UTR) ont été choisies pour les analyses phylogénétiques. Quand les séquences d'ESTs étaient trop courtes et ne contenaient pas une séquence codante complète, une séquence consensus a été créée en utilisant le logiciel Bioedit téléchargeable sur le site (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Enfin, les séquences protéiques déduites ont été vérifiées une dernière fois en interrogeant la base de données PROSITE (<http://www.expasy.org/prosite>) pour ne garder que celles portant les deux domaines LIM entiers. Les ESTs ou consensus sélectionnées pour les analyses phylogénétiques sont répertoriées en annexe 2.

### 1.3. Analyses de séquences et analyses phylogénétiques des protéines à domaine LIM

Les séquences de protéines sélectionnées ont été alignées entre elles grâce au logiciel ClustalW en utilisant les paramètres par défaut. Les éventuelles erreurs générées par le logiciel ClustalW dans l'alignement ont été vérifiées et corrigées manuellement en utilisant le logiciel Bioedit. Les analyses phylogénétiques ont été menées en utilisant les différents programmes issus du Phylogenetic Interference Package (PHYLP) 3.63 (<http://evolution.genetics.washington.edu/phylip.html>). Les matrices de distances génétiques utilisant le polymorphisme protéique ont été calculées avec le logiciel PROTDIST en utilisant la matrice JTT de substitution d'acides aminés (Jones *et al.*, 1992). Un arbre phylogénétique a

ensuite été construit en utilisant la méthode de Neighbour-joining grâce au logiciel NEIGHBOR (Saitou & Nei, 1987). Afin d'estimer la solidité statistiques des nœuds par une procédure de ré-échantillonnage avec remise fondée sur la méthode de bootstrap, 1000 échantillons bootstrap ont été générés par le logiciel SEQBOOT, et l'arbre consensus selon la règle majoritaire est créé par le logiciel CONSENSE. La famille LIM de plante a aussi été analysée à travers une méthode de parcimonie en utilisant le programme PROTPARS avec 1000 répétitions de bootstrap. Des analyses de maximum de vraisemblance ont été réalisées en utilisant le logiciel Phylml v2.4.1 accessible sur le site (<http://atgc.lirmm.fr/phylml/>) (Guindon & Gascuel, 2003) avec la matrice JTT et 100 répétitions de bootstrap. Les arbres issus des analyses de maximum de vraisemblance ont été générés avec BIONJ, un algorithme modifié de la méthode de Neighbour-joining (Gascuel, 1997). Les arbres ont été visualisés et édités avec le logiciel Treeview (Page, 1996), et les valeurs de bootstrap inférieures à 50 % n'ont pas été présentées.

Les motifs conservés dans les protéines à domaine LIM ont été détectés en utilisant l'alignement généré par ClustalW et corrigés manuellement, et en utilisant le programme MEME sur le site (<http://meme.sdsc.edu/meme/meme.html>) capable d'identifier et d'éditer des motifs de séquences consensus au sein d'un alignement de séquences protéiques (Bailey & Elkan, 1994). Les séquences protéiques alignées ont été ombrées en utilisant le logiciel Bioedit avec un seuil de 90% pour l'ombrage des résidus identiques et une matrice BLOSUM62 pour l'ombrage des résidus similaires. Le point isoélectrique (pI) et la masse moléculaire (MM) ont été prédits avec l'outil pI/Mw sur le site Internet expasy ([http://www.expasy.org/tools/pi\\_tool.html](http://www.expasy.org/tools/pi_tool.html)). Les résultats de PROSITE ont été utilisés pour trouver les sites putatifs de ASN-glycosylation et de phosphorylation.

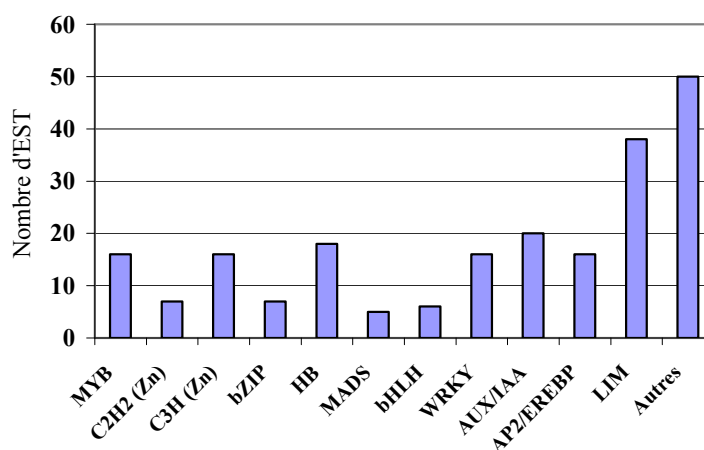
## 2 Résultats

### 2.1. Isolation des ADNc codant pour des protéines à domaine LIM chez le peuplier

La distribution des 9821 ESTs obtenues à partir de quatre banques d'ADNc différentes a montré que des ESTs homologues à la protéine à domaine LIM SF3 sont abondantes dans le xylème en différenciation de bois de tension (Baltz *et al.*, 1992b ; Déjardin *et al.*, 2004). Nos banques d'ESTs ont été obtenues à partir de différents tissus tels que la zone cambiale (ZC ; de bois tendu et opposé), le xylème en développement de bois tendu (XDT), le xylème en



développement de bois opposé (XDO), et le xylème mature (XM ; de bois tendu et opposé), prélevés sur un peuplier hybride *Populus tremula* × *P. alba* (Clone 717-1-B4) (Déjardin *et al.*, 2004). Des études bioinformatiques plus détaillées ont été réalisées pour répertorier, dans un premier temps, les ESTs homologues aux gènes codant pour des facteurs de transcription chez le peuplier. J'ai ainsi identifié au moins 125 consensus ou unigènes homologues à des facteurs de transcription dans les banques d'ADNc de l'INRA d'Orléans. On retrouve dans nos ESTs, la plupart des grandes familles de facteurs de transcription et la répartition des ESTs de facteurs de transcription entre ces différentes familles a été analysée (Fig. 32). Il n'est pas étonnant de constater que les familles de facteurs de transcription contenant le plus d'ESTs sont celles potentiellement impliquées dans le développement vasculaire (familles HD et IAA), dans la réponse à l'éthylène, une hormone synthétisée dans le bois de tension (famille AP2/EREBP), ou dans la biosynthèse des lignines (famille MYB). L'étude de la distribution des ESTs dans ces banques montre surtout que la famille de facteur de transcription la plus représentée dans ces tissus est la famille de protéines à domaine LIM avec 38 ESTs (Fig. 32).



**Figure 32. Répartition en différentes familles des EST homologues de facteurs de transcription et présentes dans les banques d'ADNc de l'INRA d'Orléans.**

Les différentes familles de facteurs de transcription sont figurées en abscisse et le nombre d'ESTs (consensus et singlet) correspondant à chaque famille est représenté en ordonné. Les familles les plus représentées dans les banques issues de zone cambiale, de jeune xylème tendu ou opposé et de xylème mature sont les familles LIM, MYB, C3H, WRKY, HB, IAA et AP2/EREBP.

J'ai donc focalisé mon étude sur la famille de protéines à domaine LIM chez le peuplier et les plantes. Cette famille est homologue à la famille CRP chez les animaux, et ces protéines comportent uniquement deux domaines LIM. Pour chaque unigène identifié codant une protéine à domaine LIM, les clones d'ADNc les plus longs ont été sélectionnés puis séquencés afin d'obtenir des séquences codantes complètes. Nos banques d'ESTs contiennent en tout six transcrits *PtaLIM* différents qui ont été nommés *PtaGLIM1a* (EF035035), *PtaGLIM1b* (EF035036), *PtaWLIM1a* (EF035038), *PtaWLIM1b* (EF035039), *PtaWLIM2a*

## Partie 1. Analyse phylogénétique

(EF035040), et *PtaPLIM2a* (EF035037). Les caractéristiques des ADNc et des protéines déduites sont figurées sur le tableau 8. Tous les ADNc pleine longueur séquencés contiennent une région codante ou « coding sequence » (CDS) entière, des régions non traduites ou « untranslated region » (UTR) en 5' et 3', et codent pour des protéines avec deux domaines LIM. Les deux domaines LIM de 52 résidus possèdent la séquence consensus suivante [C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-H] caractéristique de la famille de protéines à domaine LIM chez les plantes (Eliasson *et al.*, 2000). Il est à noter que l'ARNm *PtaPLIM2a* isolé contient un autre cadre de lecture ouvert ou « open reading frame » (ORF) codant pour une autre protéine putative qui ne présente aucune homologie avec d'autres protéines connues. La taille des protéines PtaLIM déduites varie de 189 à 206 acides aminés et la masse moléculaire calculée de ces protéines est d'environ 22 kDa. Elles ont toutes un point isoélectrique (pI) plutôt basique au alentour de 9, excepté la protéine PtaPLIM2a dont le pI (7,1) est neutre (Tableau 8).

**Tableau 8.** Caractéristiques des ADNc *PtaLIM* isolés à partir de *Populus tremula* × *P. alba*

Nom de l'ADNc et numéro d'accèsion Genbank	EST séquencée (accésion Genbank)	Taille du transcrit	Taille du 5P UTR	Taille du 3P UTR	Taille du CDS	Taille de la protéine	MM	pI
<i>PtaWLIM1a</i> EF035038	CF231651	1104	197	313	594	197	22	9,02
<i>PtaWLIM1b</i> EF035039	CF236568	945	85	266	594	197	22,1	8.95
<i>PtaGLIM1a</i> EF035035	CF235176	957	88	281	588	195	21,9	8.83
<i>PtaGLIM1b</i> EF035036	CF236473	997	119	287	591	196	22,1	8.94
<i>PtaWLIM2a</i> EF035040	CF235467	1135	300	265	570	189	20,9	9.04
<i>PtaPLIM2a</i> EF035037	CF236773	1126	332	173	621	206	22,8	7,09

Les tailles des transcrits, des régions codantes ou « coding sequence » (CDS), des régions non traduites ou « untranslated region » (UTR) sont donnés en paire de bases (pb). La taille des protéines déduites est exprimée en acide aminés (aa). MM : masse moléculaire indiquée en kDa. pI : point isoélectrique.

**Tableau 9.** Informations génomiques sur les gènes *PtLIM* de *Populus trichocarpa*.

Nom du gène	Position génomique ( <i>Populus trichocarpa</i> )	Modèle de gène <i>PtLIM</i> promu ( <i>P. trichocarpa</i> )	Nombre total d'ESTs identifiées	Taille de la protéine	MM	pI
<i>PtWLIM1a</i>	LG_XIV:598591-600291	grail3.0126006201	32	197	22	9,02
<i>PtWLIM1b</i>	LG_II:8904210-8906209	estExt_fgenesh4_pg.C_LG_II1071	20	197	22,1	8,95
<i>PtGLIM1a</i>	LG_IX:4795506-4797325	grail3.0001068602	58	195	21,9	8,83
<i>PtGLIM1b</i>	LG_I:27294474-27295832	fgenesh4_pm.C_LG_I001083	29	196	22,1	8,94
<i>PtβLIM1a</i>	LG_XII:9413188-9415996	fgenesh4_pm.C_LG_XII000273	6	191	21,4	9,11
<i>PtβLIM1b</i>	LG_XV:5205959-5209047	estExt_fgenesh4_pm.C_LG_XV0182	5	191	21,4	8,68
<i>PtWLIM2a</i>	LG_VIII:3623285-3626671	estExt_fgenesh4_pm.C_LG_VIII0239	12	189	20,9	9,04
<i>PtWLIM2b</i>	LG_X:17210641-17213976	estExt_fgenesh4_kg.C_LG_X0064	18	189	20,9	9,08
<i>PtPLIM2a</i>	LG_XIV:2231228-2232712	gw1.XIV.1147.1	5	206	22,8	7,09
<i>PtPLIM2b</i>	LG_II:11974041-11975746	grail3.0039003401	0	206	22,8	7,09
<i>PtδLIM2a</i>	LG_I:7976496-7977725	gw1.I.9051.1	0	216	24,1	8,56
<i>PtδLIM2b</i>	LG_III:11773343-11774908	gw1.III.1568.1	0	212	23,4	8,74

La taille des protéines déduites est exprimée en acide aminés (aa). MM : masse moléculaire indiquée en kDa. pI : point isoélectrique.

## 2.2. Le génome de *Populus trichocarpa* contient douze modèles de gène LIM.

La publication de la séquence du génome de *P. trichocarpa* (Tuskan *et al.*, 2006) nous a offert l'opportunité de chercher tous les modèles de gènes codant pour des protéines à domaine LIM chez le peuplier. A notre connaissance, en dehors des gènes codant pour des protéines contenant deux domaines LIM et homologues aux protéines de la famille CRP chez les animaux, il existe aussi des gènes codant pour des protéines avec uniquement un domaine LIM. Chez ces protéines, le domaine LIM est souvent lié à un autre domaine comme le domaine cytochrome P450 ou le domaine d'interaction avec l'ubiquitine. Des protéines possédant plus de deux domaines LIM répétés en tandem comme il en existe chez les animaux, n'ont pas été trouvées chez le peuplier et les autres plantes. J'ai focalisé mon étude uniquement sur les protéines contenant deux domaines LIM, habituellement appelées chez les plantes, protéines à domaine LIM ou facteur de transcription LIM. En plus des six ADNc

## Partie 1. Analyse phylogénétique

isolés dans notre laboratoire, six autres gènes codant pour des protéines à domaines LIM ont été identifiés dans le génome de *P. trichocarpa*, amenant à au moins douze le nombre de modèles de gène codant pour des protéines à domaine LIM chez le peuplier. En fonction de leurs relations phylogénétiques avec les autres protéines à domaine LIM de plante, ces gènes ont été nommés *PtWLIM1a*, *PtWLIM1b*, *PtGLIM1a*, *PtGLIM1b*, *PtβLIM1a*, *PtβLIM1b*, *PtWLIM2a*, *PtWLIM2b*, *PtPLIM2a*, *PtPLIM2b*, *PtδLIM2a* et *PtδLIM2b*. Ils sont respectivement situés sur les chromosomes 14, 2, 9, 1, 12, 15, 8, 10, 14, 2, 1 et 3 de peuplier (Tableau 9). Chez *Populus kitakamiensis*, les séquences de l'ARNm (AB079510) et du gène (AB079511) *PkLIM1*, soumis par Kumagai et Kawai en 2002 correspondent au gène *PtWLIM1a* chez *P. trichocarpa*. Nous avons observé en moyenne 98 % d'identité au niveau des séquences nucléotidiques des régions codantes ou « coding region » (CR) entre les modèles de gènes chez *P. trichocarpa* et les ADNc pleine longueur *PtaLIM* isolés chez *P. tremula* × *P. alba* (résultats non montrés). Grâce aux nouvelles informations de séquence apportées à la fois par nos ADNc *PtaLIM* pleine longueur et par les transcrits *PtLIM* isolés chez *P. trichocarpa* (Ralph *et al.*, 2006), j'ai pu corriger et améliorer l'annotation des séquences génomiques des gènes *PtLIM* de *P. trichocarpa* (Annexe 1). Les erreurs étaient surtout localisées à la jonction des introns et des exons et au niveau des UTR en 5' et 3'. Notamment, les modèles de gènes promus indiquent trois introns pour les gènes *PtδLIM2a* et *b*, alors que comme nous le verrons plus loin, les gènes *LIM* de plantes ont une structure génomique très conservée avec quatre introns. Etant donné que des séquences d'EST ou d'ADNc pleine longueur n'étaient pas disponibles pour les gènes *PtPLIM2b*, *PtδLIM2a* et *b*, nous avons cloné et séquencé les produits de RT-PCR correspondants aux transcrits des gènes *PtPLIM2b* (EF151463), *PtδLIM2a* (EF151464) et *PtδLIM2b* (EF151465) afin de déterminer les positions correctes de l'épissage des introns. Après comparaison des séquences des transcrits et génomiques, nous pouvons affirmer que les gènes *PtδLIM2a* et *b* possèdent bien quatre introns.

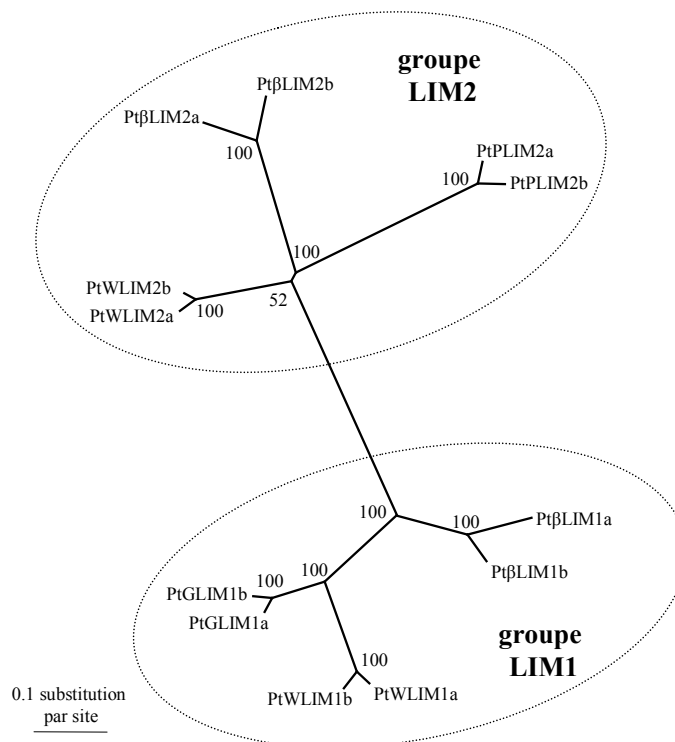
Chaque paire de gènes codant pour des protéines à domaine LIM (*PtWLIM1a* et *b*, *PtGLIM1a* et *b*, etc.) arbore une forte similarité de séquence allant de 85 % d'identité en acides aminés entre les gènes *PtδLIM2a* et *b*, jusqu'à 95 % entre les gènes *PtGLIM1a* et *b* (Tableau 10). Cette forte similarité de séquence, ainsi que les relations phylogénétiques entre les paires de protéines correspondantes, suggère fortement un événement récent de duplication qui s'est produit pour tous les gènes *LIM* chez le peuplier (Fig. 33).

## Partie 1. Analyse phylogénétique

**Tableau 10.** Comparaison paire par paire du degré d'identité en acide aminé (exprimé en pourcentage) entre les protéines à domaine LIM de peuplier.

Protein	1	2	3	4	5	6	7	8	9	10	11	12
<b>PtaWLIM1a (1)</b>	100											
<b>PtaWLIM1b (2)</b>	94	100										
<b>PtaGLIM1a (3)</b>	76	76	100									
<b>PtaGLIM1b (4)</b>	78	77	95	100								
<b>PtβLIM1a (5)</b>	64	63	65	65	100							
<b>PtβLIM1b (6)</b>	67	66	68	70	87	100						
<b>PtaWLIM2a (7)</b>	53	52	53	54	53	55	100					
<b>PtWLIM2b (8)</b>	53	52	53	54	53	55	95	100				
<b>PtaPLIM2a (9)</b>	43	44	46	46	46	49	63	62	100			
<b>PtPLIM2b (10)</b>	43	43	46	45	46	49	63	63	90	100		
<b>PtδLIM2a (11)</b>	46	46	47	48	49	49	68	68	60	61	100	
<b>PtδLIM2b (12)</b>	46	46	47	49	48	49	66	66	58	59	85	100

Pour le calcul de l'identité des paires d'acides aminés, nous avons déduit les séquences protéiques à partir des ADNc (*PtaLIM*) pleine longueur issus de *P. tremula* × *alba*, ou quand ils n'étaient pas disponibles, à partir de la séquence génomique de *P. trichocarpa* (*PtLIM*).

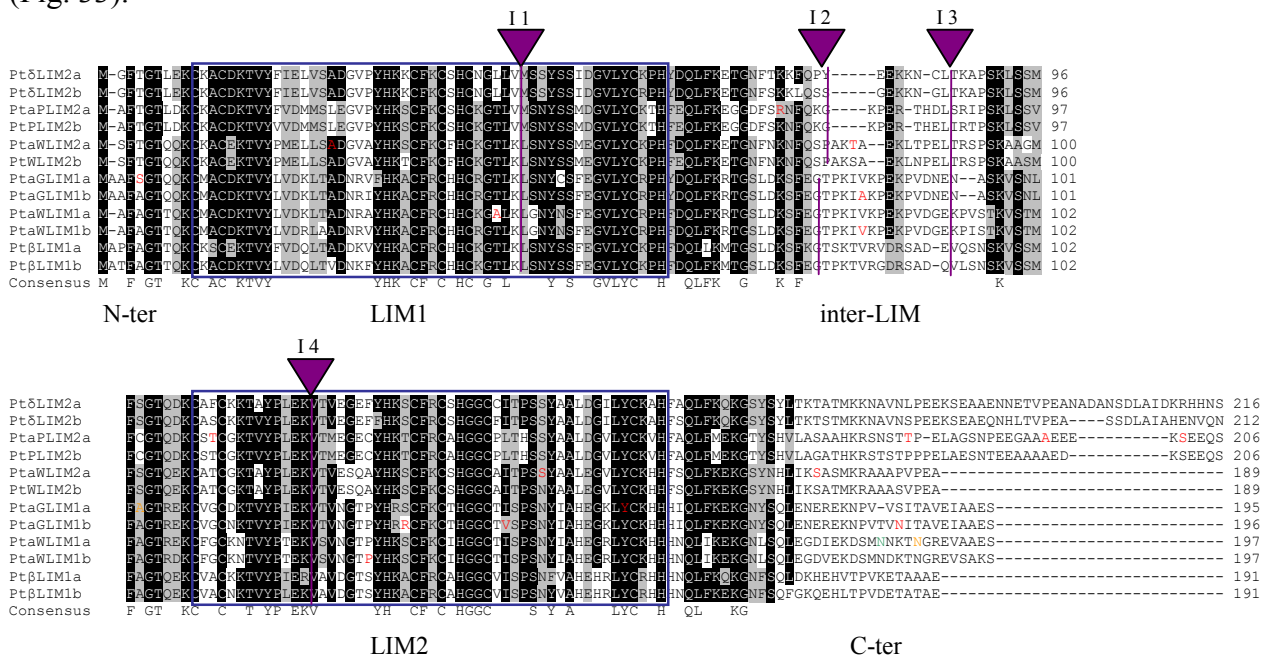


**Figure 33.** Arbre phylogénétique des protéines à domaine LIM chez *Populus trichocarpa*

Les séquences d'acides aminés des douze protéines LIM de peuplier ont été alignées au moyen du logiciel ClustalW puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des nœuds représentent des valeurs de bootstrap exprimées en pourcentage ( $\geq 50$  %) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les deux groupes LIM1 et LIM2 sont encadrés.

## Partie 1. Analyse phylogénétique

Toutes les protéines déduites ont les caractéristiques de la famille de protéines à domaine LIM chez les plantes, décrite précédemment (Eliasson *et al.*, 2000). Elles comportent deux domaines LIM de 52 acides aminés, séparés par une longue région de 40 à 47 acides aminés appelée domaine inter-LIM, un domaine N-terminal très court de 9 à 10 acides aminés, et un domaine C-terminal de longueur variable (Fig. 34). La longueur des protéines LIM est plutôt constante, entre 194 et 216 acides aminés, et leurs masses moléculaires varient entre 20,9 et 24,1 kDa (Tableau 9). Les protéines à domaine LIM de peuplier présentent un niveau de similarité de séquence allant de 43 à 95 % d'identité en acide aminé (Tableau 10). Les différences en acides aminés observées entre les protéines à domaine LIM de peuplier sont surtout localisées dans le domaine inter-LIM et le domaine C-terminal. À partir de la seule analyse d'identité en acides aminés entre les différentes protéines à domaine LIM de peuplier, nous pouvons différencier deux groupes majeurs LIM1 et LIM2. Au sein de chaque groupe, le pourcentage d'identité en acide aminé entre les protéines est plus fort (de 57 à 95 %) qu'entre les protéines des deux groupes (de 43 à 53 %). L'arbre phylogénétique de la famille de protéines à domaine LIM chez le peuplier confirme l'existence de ces deux groupes (Fig. 33).



**Figure 34.** Alignement des protéines à domaine LIM de peuplier.

Les acides aminés des douze protéines LIM de peuplier ont été alignés en utilisant le logiciel ClustalW. L'alignement a été édité avec le logiciel Bioedit. La séquence consensus avec un seuil de 90 % est montrée en bas des séquences alignées. Les gaps sont marqués par des tirets, les résidus conservés et identiques sont mis en évidence par un ombrage noir, et les résidus similaires sont ombrés en gris selon la matrice BLOSUM62. Les domaines LIM sont marqués par des boîtes rouges et la position des introns par des lignes verticales et des triangles violets. Les substitutions d'acides aminés entre *Populus trichocarpa* (Pt) et *Populus tremula* × *P. alba* sont indiquées par des lettres rouges, celles au sein de *Populus tremula* × *P. alba* sont indiquées par des lettres oranges et celle entre *Populus trichocarpa* et *P. kitakamiensis* par une lettre verte.

**Tableau 11.** Polymorphismes simples de nucléotide (SNPs) identifiés au sein de la population d'ESTs issue de peuplier hybride *Populus tremula* × *P. alba*, et SNPs observés entre *Populus tremula* × *P. alba* et *P. trichocarpa*.

Nom de l'ADNc et numéro d'accèsion Genbank	Consensus ID	Nombre d'ESTs trouvées	Positions des SNPs <i>Populus tremula</i> × <i>alba</i>	Changement en acide aminé	Positions des SNPs <i>Populus trichocarpa</i>	Changement en acide aminé
<i>PtaWLIM1a</i> EF035038	cn1788	6	127 (G → A) 566 (A → G)	43 (A → T) 189 (N → S)	127 (G → A)	43 (A → T)
<i>PtaWLIM1b</i> EF035039	cn1789	7	Non trouvé		247 (G → T) 385 (C → T)	83 (V → L) 129 (P → S)
<i>PtaGLIM1a</i> EF035035	cn1720 et cn1721	15	308 (C → G)	103 (A → G)	13 (T → G) 446 (T → C)	5 (S → A) 156 (Y → H)
<i>PtaGLIM1b</i> EF035036	cn1719	8	Non trouvé		251 (C → T) 396 (A → C) 427 (G → A) 557 (A → G)	84 (A → V) 132 (R → S) 143 (V → I) 186 (N → S)
<i>PtaWLIM2a</i> EF035040	Singlet	1	Non trouvé		73 (G → A) 244 (A → T) 437 (G → A) 527 (C → G)	25 (A → T) 82 (T → S) 146 (S → N) 176 (S → C)
<i>PtaPLIM2a</i> EF035037	Singlet	1	Non trouvé		221 (T → C) 319 (A → G) 547 (A → C) 590 (C → T) 604 (T → C)	74 (R → K) 107 (T → A) 183 (T → P) 197 (A → V) 202 (S → P)

Les positions des polymorphismes simples de nucléotide (SNPs) observés dans les populations d'ESTs de *Populus tremula* × *alba* sont calculées à partir du CDS et données en paire de bases. Les SNPs observés chez *P. trichocarpa* par comparaison avec les séquences d'ADNc pleine longueur *PtaLIM* issus de *Populus tremula* × *alba* sont également indiqués. Les positions des changements consécutifs en acides aminés dans les protéines déduites sont données en nombre d'acides aminés.

### 2.3. Identification de polymorphisme de séquences nucléotidique et protéique au sein des espèces de peuplier

Quand les séquences nucléotidiques des ESTs sont examinées au sein de chaque consensus *PtaLIM*, deux populations différentes d'ESTs sont trouvées pour les gènes *PtaGLIM1a* et *PtaWLIM1a*. Cette observation est peut-être la conséquence de la présence de deux formes alléliques du gène au sein du consensus, l'une provenant du parent *Populus tremula* et l'autre provenant du parent *Populus alba*. Ces formes alléliques contiennent des substitutions de nucléotides qui conduisent dans certain cas à des substitutions d'acides aminés (Tableau 11). La plupart du temps, ces polymorphismes simples de nucléotide ou (SNP) pour « Single Nucleotide Polymorphisms » génèrent des substitutions d'acides aminés ayant des fonctions biochimiques conservées. Il existe aussi des SNPs silencieux non indiqués sur le tableau 11 qui touchent la troisième base du codon et n'introduisent pas de changement

## Partie 1. Analyse phylogénétique

en acide aminé. De la même manière, comparés à nos ADNc isolés au laboratoire, les gènes de *Populus trichocarpa* contiennent aussi des SNP introduisant des substitutions d'acides aminés dans les protéines à domaine LIM. Nous pouvons observer en moyenne trois substitutions d'acides aminés entre les protéines déduites de *P. trichocarpa* et celles déduites de *P. tremula* × *P. alba*. Par exemple, comparée à la séquence de l'ADNc soumise à la base de donnée Genbank, la seconde version allélique du gène *PtaWLIM1a* au sein du consensus cn1788 et la séquence génomique de *PtWLIM1a* contiennent toutes les deux un polymorphisme de nucléotide simple à la position 127 de la séquence codante conduisant à une substitution d'acide aminé majeure à la position 43. En effet, à cette position, l'acide aminé alanine (A) présent dans la protéine déduite de l'ADNc *PtaWLIM1a* (obtenu chez *P. tremula* × *P. alba*) est remplacé par un acide aminé thréonine (T) participant à un site putatif de phosphorylation dans la séquence de la protéine PtWLIM1a déduite du génome de *P. trichocarpa*. La protéine déduite PkLIM1 issue de *P. kitakamiensis* contient aussi une substitution d'acide aminé à la position 185, où une sérine remplace l'asparagine présente dans protéine PtWLIM1a déduite du génome de *P. trichocarpa*. Le remplacement de cette asparagine par une sérine introduit un site putatif de phosphorylation supplémentaire dans la protéine PkLIM1. Pour les gènes *PtaGLIM1b* et *PtaWLIM1b*, aucune autre version allélique n'a été identifiée au sein des populations d'ESTs disponibles au laboratoire. Les transcrits *PtaWLIM2a* et *PtaPLIM2a* proviennent de singlets et nous sommes donc incapable d'identifier des allèles pour ces deux gènes. Par contre pour ces quatre gènes *PtaGLIM1b*, *PtaWLIM1b*, *PtaWLIM2a* et *PtaPLIM2a*, des SNPs induisant des substitutions d'acides aminés ont été observés entre les ADNc soumis et la séquence génomique de *P. trichocarpa* (Tableau 11). On peut remarquer sur ce tableau que les protéines PtaGLIM1b, PtaWLIM2a et PtaPLIM2a contiennent en moyenne deux fois plus de substitutions en acide aminé que les protéines PtaGLIM1a, et PtaWLIM1a et b. Enfin, d'autres différences alléliques ont été détectées entre les ADNc *PtaLIM* isolés au laboratoire et les séquences d'ESTs provenant d'autres taxons de peuplier comme *P. deltoïdes*, *P. tremula* × *tremuloïdes*, *P. euphratica*, *P. canadensis* et *P. tomentiglandelosa* (résultats non montrés).

De façon tout aussi intéressante pour la recherche de marqueurs génétiques, plusieurs motifs microsatellites ont été détectés dans les régions non traduites (UTR) de certains gènes *PtLIM*. Pour le gène *PtGLIM1a* chez *P. trichocarpa*, un motif microsatellite « CTT » répété six fois est situé dans le 5' UTR proche du codon d'initiation de la traduction ATG (Annexe 1). Ce motif est répété cinq fois dans l'ADNc *PtaGLIM1a* isolé à partir de *Populus tremula* ×



## Partie 1. Analyse phylogénétique

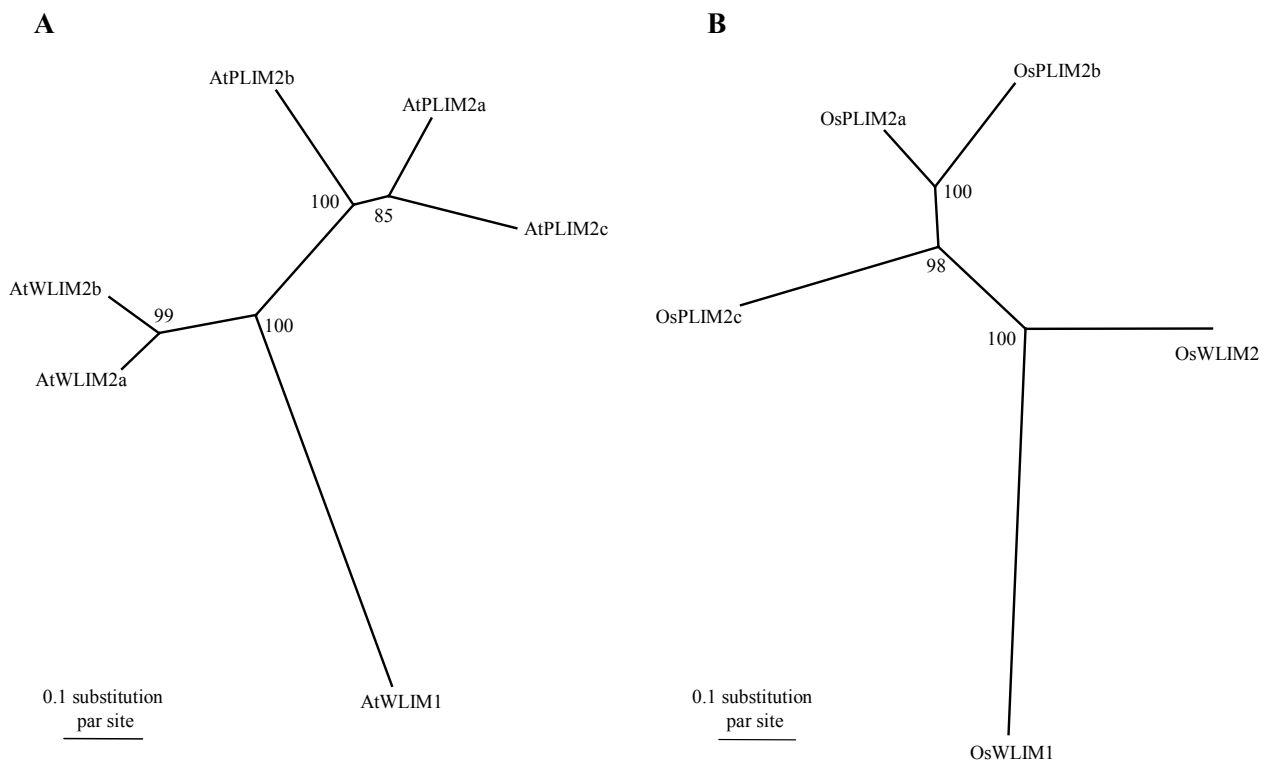
*P. alba*. Le gène *PtβLIM1a* contient un motif microsatellite « AT » répété treize fois dans son 3' UTR. Enfin le gène *PtβLIM1b* contient dans son 5' UTR un motif microsatellite « GAA » répété quatre fois suivi d'un autre motif microsatellite « ATAG » aussi répété quatre fois.

### 2.4. Les génomes d'*Arabidopsis* et de riz contiennent six modèles de gènes LIM

Une étude précédente a rapporté l'identification de trois gènes LIM dans le génome d'*Arabidopsis thaliana* (l'arabette des dames) : *AtWLIM1* (X74319), *AtPLIM2* (AC004665), et *AtWLIM2* (AF002109) (Eliasson *et al.*, 2000). Étant donné que les séquences des génomes d'*Arabidopsis* et de riz sont maintenant disponibles au public, j'ai eu l'opportunité d'identifier de manière approfondie tous les gènes codant pour des protéines à domaine LIM. Les analyses bioinformatiques menées au sein des bases de données NCBI et TIGR montrent que le génome d'*Arabidopsis* contient trois autres gènes *AtLIM* (Fig. 35A). Parce que ces gènes semblent être dupliqués, je les ai nommés *AtPLIM2b*, *AtPLIM2c* et *AtWLIM2b* aussi appelé *AtL2* (Tranbarger *et al.*, 2003). Ainsi les précédents gènes *AtPLIM2* et *AtWLIM2* ont été renommés *AtPLIM2a* et *AtWLIM2a* respectivement. Le tableau 12 liste les numéros d'accessions des clones génomiques et d'ADNc pour les six gènes LIM d'*Arabidopsis*, ainsi que les clones d'ADNc ou génomiques contenant des erreurs ou codant pour des protéines LIM partielles. Par exemple, pour le gène *AtWLIM1*, le clone génomique AE005172 doit être considéré à la place du précédent clone identifié (X74319) où le premier intron est reconnu à tort comme une région codante. Par ailleurs, le gène *AtWLIM1* est représenté par deux ADNc AF370569 et AY085377. Les gènes *AtWLIM2a* et *AtWLIM2b* sont des gènes étroitement similaires, probablement dupliqués et respectivement représentés par les ADNc AY094448 et AY084361. Par le nombre élevé d'ESTs trouvées dans les bases de données publiques pour le gène *AtWLIM2b*, nous pouvons supposer qu'il est plus exprimé chez *Arabidopsis* que le gène *AtWLIM2a*. Pour les trois gènes *AtPLIM2a*, *b*, et *c*, le nombre d'ESTs indiqué dans les bases de données NCBI et TIGR est surestimé. Ces estimations incluent des ESTs homologues à des gènes physiquement proches sur le génome. En effet, les gènes *AtPLIM2a*, *b* et *c* sont respectivement chevauchants avec un gène codant pour une phosphomannomutase (At2g45790), un gène codant pour une protéine inconnue (At1g01770), et un gène codant pour une oxydoréductase (At3g61220). Le tableau 12 indique le nombre exact d'ESTs répertoriées correspondants effectivement à ces trois gènes *AtPLIM2a*, *b* et *c*. Le gène *AtPLIM2a* est soutenu par les ADNc AK118044 et BT005586. On peut noter que pour ce

## Partie 1. Analyse phylogénétique

gène, un variant transcriptionnel contenant le premier intron non épissé est indiqué dans la base de donnée du NCBI. Ce variant n'est représenté que par une seule EST parmi les 13 identifiées et code pour une protéine contenant uniquement le deuxième domaine LIM. Concernant le gène *AtPLIM2b*, correctement représenté par l'ADNc BX817923, il existe un autre ADNc « chimère » (AY059790) qui, en plus du cadre de lecture ouvert correspondant au gène *AtPLIM2b*, porte un second cadre de lecture ouvert en 5' correspondant au gène *At1g01770*. Le gène *AtPLIM2c* est représenté par l'ADNc BT022103 et une seule EST.



**Figure 35.** Arbres phylogénétiques des protéines à domaine LIM chez *Arabidopsis thaliana* et *Oryza sativa*. Les deux arbres phylogénétiques ont été construits à partir des six protéines à domaine LIM d'*Arabidopsis* (A) et des six protéines à domaine LIM de riz (B). Les séquences d'acides aminés ont été alignées au moyen du logiciel ClustalW puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap exprimées en pourcentage ( $\geq 50\%$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre.

Partie 1. Analyse phylogénétique

**Tableau 12.** Informations génomiques sur les gènes *LIM* chez *Arabidopsis thaliana* (At) et *Oryza sativa* (Os)

Nom du gène	Position génomique	N° d'accèsion génomique	Clone génomique BAC/PAC	N° d'accèsion des ADNc validés et sélectionnés	N° d'accèsion des ADNc ou clones génomiques incomplets ou contenant des erreurs.	Nombre d'ESTs trouvées
<i>AtWLIM1</i>	At1g10200	AE005172	F14N23.8	AF370569, AY085377	AY072143	75
<i>AtWLIM2a</i>	At2g39900	AF002109	T28M21.6	AY094448, AY122901	BX842053, BX842049	23
<i>AtWLIM2b</i>	At3g55770	AL161667	F1I16.180	AY084361, AY063924, AY091249, BX822878, BX824310	AF370574, X91398, BX822901, BX842219	159
<i>AtPLIM2a</i>	At2g45800	AC004665	F4I18.22	AK118044, BT005586	BX821642	13
<i>AtPLIM2b</i>	At1g01780	AE005172	T1N6.19	BX817923	BX817691, AY059790	15
<i>AtPLIM2c</i>	At3g61230	AL137898	T20K12.130	BT022103	BX825681	1
<i>OsWLIM1</i>	Os12g32620	AL837527	OSJNBb0078B20	Non trouvé	AK058220	19
<i>OsLIM</i>	Os06g13030	AP004995	OSJNBa0068B06	AK102383	Non trouvé	22
<i>OsWLIM2</i>	Os03g15940	AC135208, AC135205	OJ1364E02, OJ1012B02	AK058507	Non trouvé	92
<i>OsPLIM2a</i>	Os02g42820	AP004133, AP004768	OJ1112_G03, P0010C01	Non trouvé	Non trouvé	2
<i>OsPLIM2b</i>	Os04g45010	AL662984, AL663006	OSJNBa0081C01, OSJNBb0039L24	AK069907	Non trouvé	23
<i>OsPLIM2c</i>	Os10g35930	AC078840, AE016959	OSJNBb0073N24	AK072520	Non trouvé	59

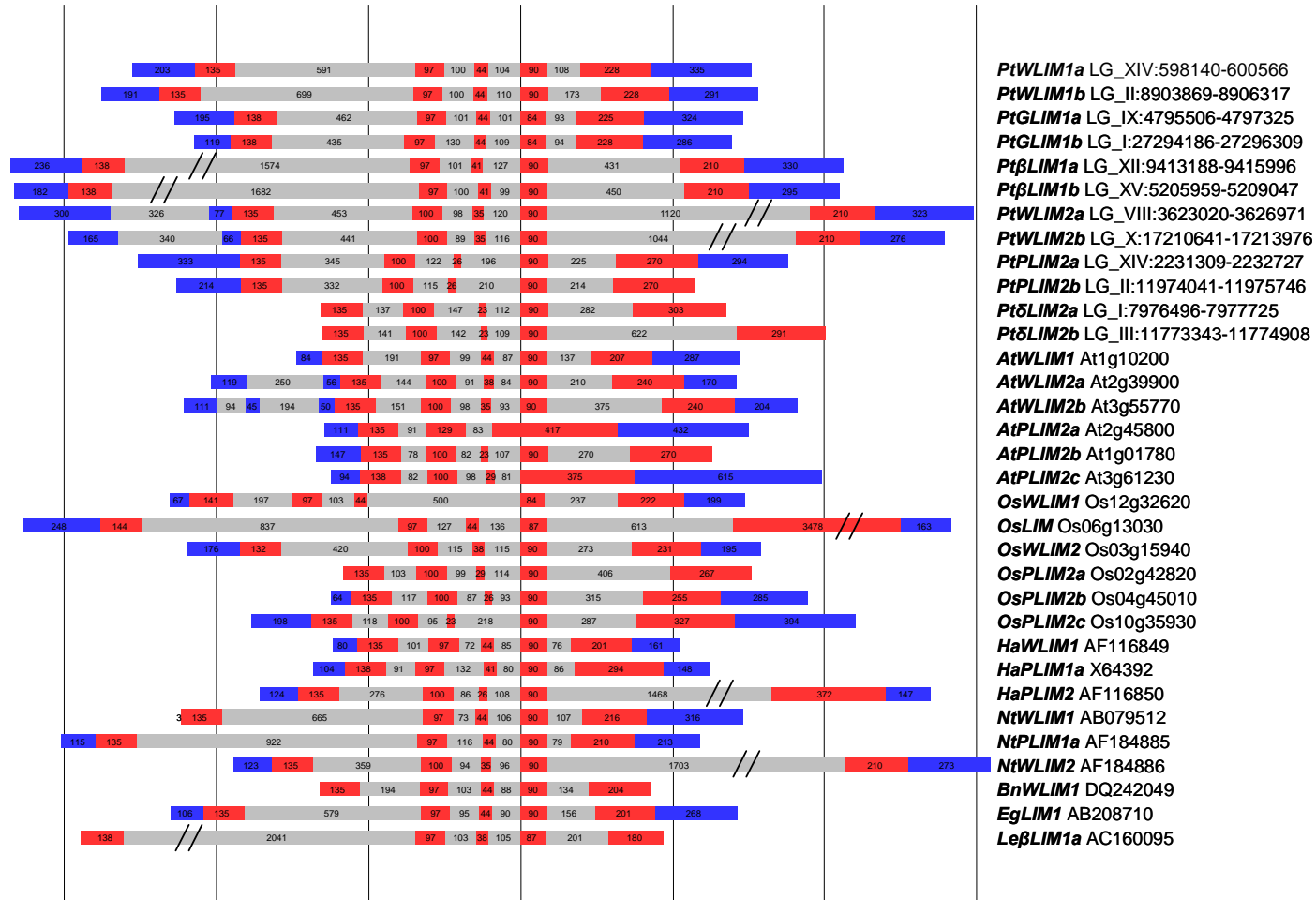
## Partie 1. Analyse phylogénétique

Comme chez *Arabidopsis*, le génome du riz (*Oryza sativa*) contient six gènes codant pour des protéines à domaine LIM : *OsWLIM1*, *OsWLIM2*, *OsPLIM2a*, *OsPLIM2b*, *OsPLIM2c* et *OsLIM* (Tableau 12 et Fig. 35B). *OsPLIM2a*, *b* et *c* sont très similaires dans leurs séquences et peuvent donc être considérés comme des gènes dupliqués ou paralogues. Contrairement aux gènes *LIM* de peuplier tous dupliqués, seuls certains gènes *LIM* chez *Arabidopsis* et le riz apparaissent dupliqués. Comme pour *Arabidopsis thaliana*, le tableau 12 récapitule les informations génomiques et transcriptomiques disponibles pour ces gènes. Les gènes *LIM* de riz sont localisés sur différents chromosomes et sont tous portés par des clones BAC ou PAC. Les gènes *OsWLIM2*, *OsPLIM2b* et *OsPLIM2c* sont respectivement représentés par les ADNc AK058507, AK069907 et AK072520 (Kikuchi *et al.*, 2003), ainsi que par un nombre important d'ESTs. Pour le gène *OsWLIM1*, l'ADNc identifié (AK058220) est tronqué et aucune EST codant pour une protéine entière n'a pu être récupérée. Le gène *OsPLIM2a* est aussi faiblement représenté par des séquences transcrites, avec seulement deux ESTs trouvées et aucun ADNc pleine longueur publié. C'est pourquoi, pour ces deux gènes, j'ai utilisé les protéines déduites des séquences génomiques pour l'alignement des séquences lors des analyses phylogénétiques. Dans le cas du gène *OsLIM*, j'ai identifié un transcrit (AK102383) très long de 4323 paires de bases codant pour une protéine à deux domaine LIM inattendue de 1303 acides aminés. Les deux domaines LIM classiques sont suivis d'un domaine C-terminal très long comportant une région riche en acide glutamique et n'ayant aucune homologie avec des protéines connues. Seulement une EST (CI584223) parmi les 22 ESTs trouvées par une recherche BlastN est localisée dans la partie 5' du transcrit au niveau du premier domaine LIM. Du fait de ce long domaine C-terminal inhabituel, je n'ai pas inclus cette séquence *OsLIM* dans les analyses phylogénétiques.

### 2.5. Organisation génomique des gènes *LIM* chez le peuplier, *Arabidopsis* et le riz

A partir de l'analyse génomique, nous pouvons affirmer que, excepté les gènes *AtPLIM2a* et *c* possédant respectivement deux et trois introns, tous les gènes *LIM* de plantes ont une structure de gène très conservée avec quatre introns situés au sein de la séquence codante (Fig. 36). Chez le peuplier et *Arabidopsis*, les gènes *WLIM2* diffèrent des autres gènes *LIM* par la présence d'un ou de deux introns supplémentaires dans la région non traduite en 5'.

## Partie 1. Analyse phylogénétique



**Figure 36.** Organisation génomique des gènes *LIM* de plante.

La structure introns/exons des séquences génomiques est représentée pour les familles de gènes *LIM* chez *Populus trichocarpa* (Pt), *Arabidopsis thaliana* (At) et *Oryza sativa* (Os). D'autres gènes *LIM* de plantes issus des espèces *Helianthus annuus* (Ha), *Nicotiana tabacum* (Nt), *Eucalyptus globulus* (Eg), *Lycopersicon esculentum* (Le), et *Brassica napus* (Bn) sont aussi indiqués. Les régions codantes (CR), les introns, et les régions non traduites en 5' et 3' (UTR) sont respectivement représentés par des rectangles rouges, gris clair, et bleu. A l'intérieur de chaque rectangle, les longueurs respectives des CR, introns, et des 5' et 3' UTR sont données en paires de base (pb). Du fait de la limitation de place, les régions codantes et les introns trop longs sont interrompus par des barres double. La distance entre chaque grande barre verticale est de 500 paires de base. Les positions ou les clones génomiques sont indiqués à la droite de chaque gène.

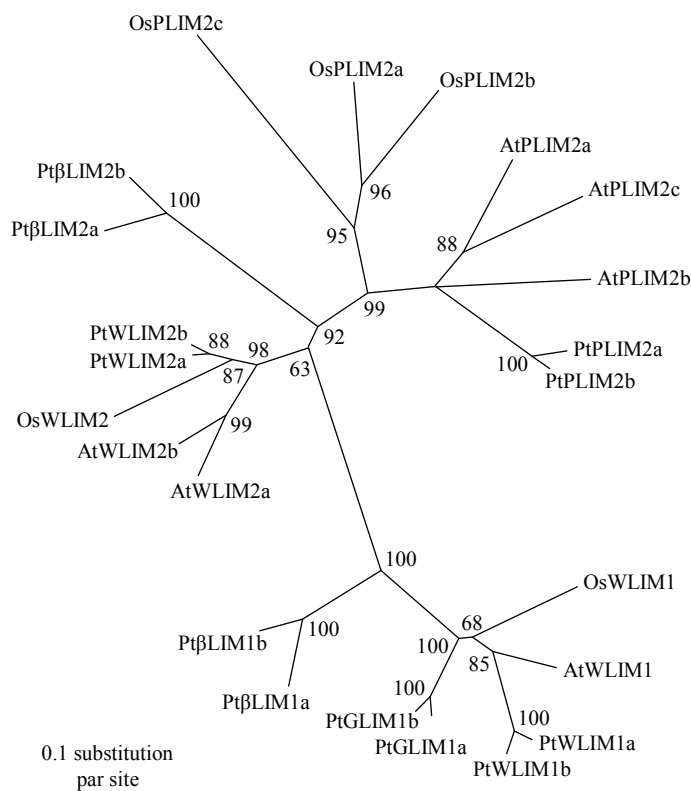
## Partie 1. Analyse phylogénétique

Grâce aux données génomiques et contrairement à la précédente caractérisation du gène *AtWLIM2a* (AF002109) décrivant quatre introns (Eliasson *et al.*, 2000), j'ai détecté un cinquième intron de 250 paires de base, et localisé 56 paires de base avant le codon ATG d'initiation de la traduction. Le gène dupliqué *AtWLIM2b* contient six introns. Les deux introns supplémentaires de 94 et 194 paires de base respectivement sont aussi localisés dans la région non traduite en 5'. Quatre épissages alternatifs du gène *AtWLIM2b* sont indiqués dans les bases de données NCBI et TIGR. Ces variants correspondent tous à une variation dans l'épissage du premier intron sans aucune conséquence sur l'intégrité de la protéine. Toutefois, les analyses BlastN montrent que ces variants transcriptionnels ne sont représentés que par une ou deux ESTs parmi les 164 ESTs identifiées pour ce gène. Les introns supplémentaires des gènes *PtWLIM2a* et *b* de peuplier ont respectivement une longueur de 326 et 340 paires de base, et sont respectivement localisés, dans la région 5' non traduite à 77 et 66 paires de bases avant le codon ATG d'initiation de la traduction. On peut noter que l'on trouve trois délétions de six à huit nucléotides dans la région 5' non traduite du transcrit *PtaWLIM2a* comparé à la séquence génomique *PtWLIM2a*. Comme nous le verrons plus tard, ces délétions peuvent correspondre à des variations d'épissage de ce cinquième intron situé dans le 5' UTR. Une structure intronique dans la région 5' non traduite du gène *NtWLIM2* n'a pas pu être prédite du fait que la séquence génomique disponible (AF184886) commence au codon ATG d'initiation de la traduction.

La longueur des régions codantes CR1, CR2 et CR4 est très conservée entre les gènes *LIM* de plante (Fig. 36). La première région codante (CR1) est longue de 135 à 138 paires de base. Seuls les gènes *OsWLIM1*, *OsWLIM2* et *OsLIM* ont une CR1 de longueur variable. Les positions du premier et du dernier intron sont très conservées dans les premier et second domaines LIM respectivement (Fig. 34). Ainsi, seules les différences observées dans la longueur de la première région codante reflètent les variations observées dans la longueur du domaine N-terminal. La seconde région codante est longue de 97 paires de bases pour les gènes appartenant au groupe LIM1, et est longue de 100 paires de base pour les gènes appartenant au groupe LIM2. La troisième région codante est très courte et sa longueur varie de 23 à 44 paires de bases entre les gènes *LIM*. Cette variation de la longueur de la CR3 souligne les différences observées entre les protéines, dans le nombre et la composition en acides aminés du domaine inter-LIM connectant les deux domaines LIM. Nous pouvons remarquer que les gènes *PtGLIM1a* et *b* et les gènes *PtWLIM1a* et *b* ont une CR3 de même longueur (44 pb). La quatrième région codante (CR4), localisée dans le second domaine LIM, est la région codante la plus conservée avec une longueur de 90 paires de base pour la

## Partie 1. Analyse phylogénétique

majorité des gènes *LIM* de plante. Seul les gènes *PtGLIM1a* et *b* et *OsWLIM1* vont à l'encontre de cette règle avec une CR4 longue de 84 paires de base. La longueur de la cinquième région codante est très variable en longueur pour tous les gènes *LIM* de plante reflétant les différences observées dans la longueur du domaine C-terminal des protéines à domaine LIM. Excepté les gènes *PtGLIM1a* et *b* et *Pt $\delta$ LIM2a* et *b*, les gènes dupliqués de peuplier ont généralement une région codante CR5 de même longueur. Enfin, comparé aux autres gènes *LIM* de plante, les gènes *PLIM2* et  *$\delta$ LIM2* possèdent la CR5 la plus longue et codent ainsi pour des protéines ayant un domaine C-terminal extrêmement long. La longueur des introns est relativement variable entre les gènes *LIM*, et varie même entre les paires de gènes *LIM* dupliqués chez le peuplier. Les caractéristiques les plus marquantes sont les longueurs importantes du premier intron des gènes *Pt $\beta$ LIM1a* et *b* (environ 1500 pb) et du dernier intron des gènes *PtWLIM2a* et *b* (environ 1100 pb).



**Figure 37.** Arbre phylogénétique des protéines à domaine LIM chez *Populus trichocarpa*, *Arabidopsis thaliana* et *Oryza sativa*

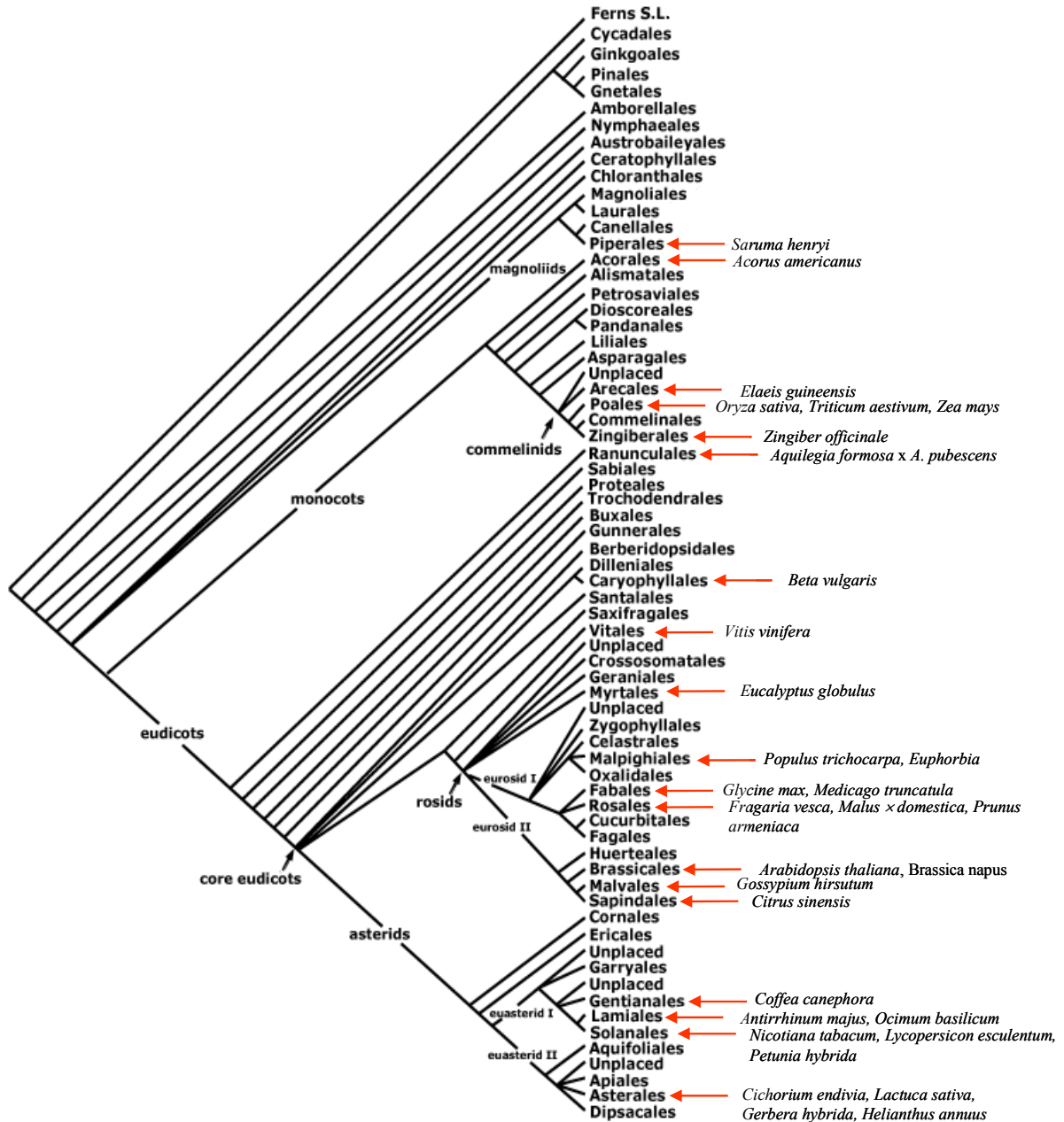
Les séquences d'acides aminés des 24 protéines LIM de peuplier d'*Arabidopsis* et de riz ont été alignées au moyen du logiciel ClustalW puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des nœuds représentent des valeurs de bootstrap exprimées en pourcentage ( $\geq 50$  %) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM : At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Pt, *Populus trichocarpa*.

### 2.6. Identification d'étiquettes de séquences exprimées (ESTs) homologues aux protéines à domaine LIM chez les plantes

Dans une précédente étude, les protéines à domaine LIM de tournesol (*Helianthus annuus*), de tabac (*Nicotiana tabacum*) et d'*Arabidopsis* ont été classées dans deux groupes LIM1 et LIM2 et sub-divisées en quatre sous-groupes : les sous-groupes PLIM1 et PLIM2 spécifiquement exprimés dans le pollen et les sous-groupe WLIM1 et WLIM2 exprimés dans toute la plante (Eliasson *et al.*, 2000). Grâce à la disponibilité des génomes séquencés de *Populus trichocarpa*, *Arabidopsis thaliana* et *Oryza sativa*, j'ai identifié un nombre plus important de gènes appartenant à la famille de protéine à domaine LIM chez les plantes (Fig. 37). Les nouvelles protéines à domaine LIM découvertes peuvent définir de nouveaux groupes ou sous-groupes, ou être apparentées aux sous-groupes identifiés auparavant. Afin d'approcher la diversité des protéines à domaine LIM chez les plantes, j'ai effectué une recherche extensive des ADNc et ESTs codant pour ces protéines à deux domaine LIM dans la base de donnée NCBI de séquences de plantes, en utilisant l'outil BlastN. En dehors des gènes de peuplier, de riz et d'*Arabidopsis*, j'ai trouvé 206 unigènes codant pour des protéines homologues aux protéines à domaine LIM de plante. Par contre, je n'ai pas inclus 51 d'entre eux dans les analyses phylogénétiques car ces unigènes contenaient des séquences codantes partielles (résultats non montrés). Parmi les 155 unigènes codant pour une protéine LIM entière, 125 ont une EST, une séquence codante (CDS) ou un ADNc représentatif, alors que pour les 30 unigènes restants un contig a été créé pour construire une séquence consensus portant une séquence codante complète (Annexe 2). Ces 155 unigènes codant pour une protéine LIM entière ont été obtenus à partir de 2228 ESTs et ADNc collectés chez différentes plantes. Nous observons une moyenne d'environ 14 ESTs par unigène sachant que certains unigènes ne sont représentés que par une EST, et que le maximum d'ESTs trouvées par unigène est de 201 (unigène *ZmPLIM2a* chez le maïs). Mis à part les 24 gènes d'*Arabidopsis thaliana*, d'*Oryza sativa* et de *Populus trichocarpa*, j'ai identifié des ESTs provenant de 62 espèces différentes issues de différentes divisions de plantes : Les bryophytes ou mousses, les ptéridophytes ou fougères, les gymnospermes ou conifères, et les angiospermes. Selon la nouvelle nomenclature taxonomique, les ptéridophytes, les monocotylédones, et les gymnospermes sont maintenant respectivement appelés les filicopsidées, les liliopsidées, et les coniféropsidées (The Angiosperm Phylogeny Group, 2003).



## Partie 1. Analyse phylogénétique



**Figure 38.** Arbre phylogénétique des différents ordres chez les angiospermes  
Les espèces de plantes, pour lesquelles des ESTs ont été isolées (indiquées par des flèches rouges), ont été placées au sein des différents ordres de l'arbre phylogénétique des angiospermes. Cet arbre est issu du site Internet (<http://www.mobot.org/MOBOT/Research/APweb/welcome.html>)

Chez les angiospermes, sont représentés les classes des monocotylédones ou graminées, et des eudicotylédones (ou vraie dicotylédones) incluant les magnoliopsidées et les core eudicotylédones (ou noyau des vraie dicotylédones) (Fig. 38 et Annexe 3). Au sein de la large classe des eudicotylédones, des gènes *LIM* provenant d'espèces appartenant aux sous-classes des rosidées et des astéridées, ainsi qu'appartenant aux ordres des *Caryophyllales* et *Piperales* ont été identifiés. Chez le tabac et le tournesol, le nombre de gène *LIM* a augmenté

## Partie 1. Analyse phylogénétique

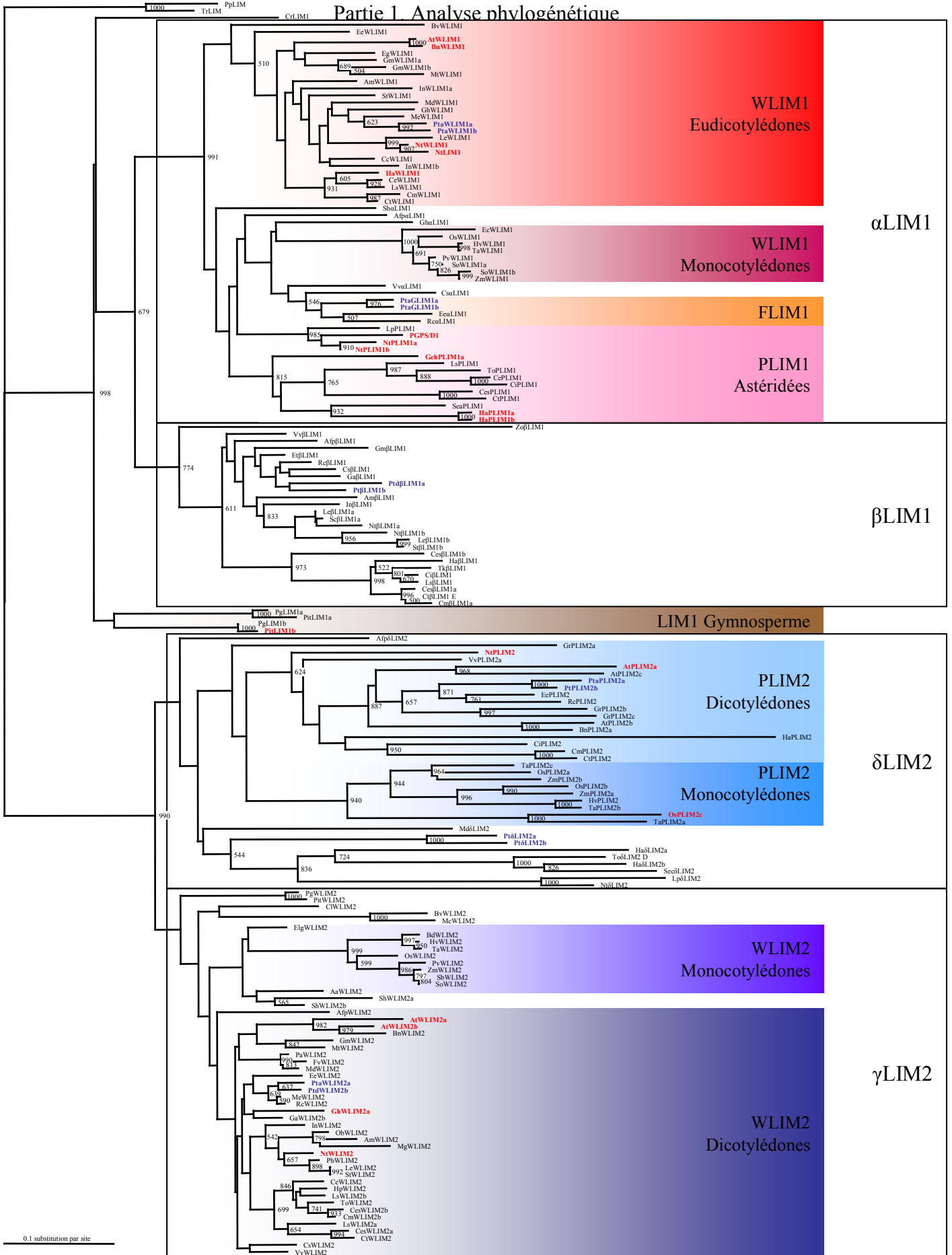
grâce à la découverte de nouvelles ESTs, notamment les deux ESTs *HpWLIM2* issues d'*Helianthus petiolaris*, dont l'existence du gène avait été précédemment suspectée dans ce genre (Eliasson *et al.*, 2000). Au total, pour ces deux espèces, huit transcrits *NtLIM* ou *HaLIM* différents ont ainsi été identifiés.

### 2.7. Analyse phylogénétique des protéines à domaine LIM de plante

#### 2.7.1. Quatre groupes différents $\alpha$ LIM1, $\beta$ LIM1, $\gamma$ LIM2, et $\delta$ LIM2

Les arbres phylogénétiques ont été construits avec les séquences d'acides aminés déduites des gènes codant pour des protéines à domaine LIM. J'ai renommé les protéines à domaine LIM de plante en fonction de leurs relations phylogénétiques. Les groupes LIM1 et LIM2 précédemment identifiés (Eliasson *et al.*, 2000) sont clairement séparés et leur séparation est confortée par une forte valeur de bootstrap au niveau des protéines PpLIM et TrLIM identifiées chez les mousses *Physcomitrella patens* et *Tortula ruralis* (Fig. 39). Nous avons été incapables de placer ces deux protéines à domaine LIM dans l'un des deux groupes LIM1 et LIM2 car leurs séquences montrent à la fois des similarités pour ces deux groupes. Ainsi, les arbres phylogénétiques ont été enracinés en utilisant les séquences de ces deux protéines à domaine LIM de bryophytes. La famille de protéines à domaine LIM de plante peut être divisée en quatre groupes  $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2, et  $\delta$ LIM2, résultant de la division des deux groupes LIM1 et LIM2. Les longueurs des branches et les fortes valeurs de bootstraps montrent que ces quatre groupes sont phylogénétiquement bien séparés. Par ailleurs, cette analyse phylogénétique confirme l'existence des sous-groupes PLIM1, WLIM1, PLIM2 et WLIM2 précédemment décrits (Eliasson *et al.*, 2000). Le groupe  $\beta$ LIM1 est un nouveau groupe qui n'avait pas été identifié auparavant, alors que le groupe  $\alpha$ LIM1 comprend les sous groupes PLIM1 et WLIM1. Les sous-groupes WLIM2 et PLIM2 appartiennent respectivement aux groupes  $\gamma$ LIM2, et  $\delta$ LIM2. De plus, chaque groupe contient de nouveaux sous-groupes qui seront décrit ci-dessous.

# Partie 1 Analyse phylogénétique



## Partie 1. Analyse phylogénétique

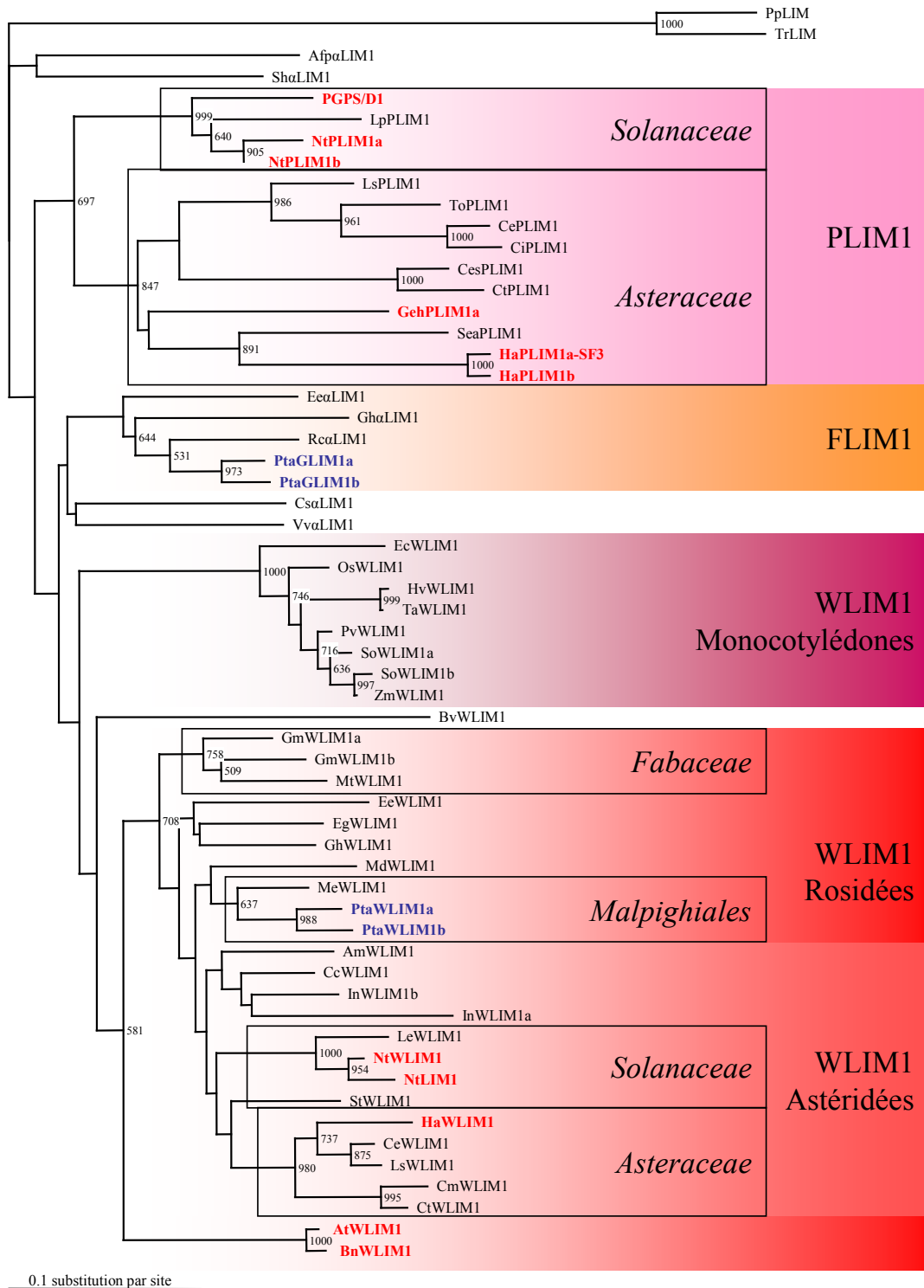
**Figure 39.** Arbre phylogénétique de la famille de protéines à domaine LIM chez les plantes.

Les séquences d'acides aminés de 178 protéines à domaine LIM ont été alignées au moyen du logiciel ClustalW, puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leur appartenance aux groupes ou sous-groupes WLIM1, PLIM1, FLIM1,  $\alpha$ LIM1,  $\beta$ LIM1, WLIM2, PLIM2 et  $\delta$ LIM2 qui sont encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM : Aa, *Acorus americanus*; Afp, *Aquilegia formosa* x *A. pubescens*; Am, *Antirrhinum majus*; At, *Arabidopsis thaliana*; Bd, *Brachypodium distachyon*; Bn, *Brassica napus*; Bv, *Beta vulgaris*; Cc, *Coffea canephora*; Ce, *Cichorium endivia*; Ci, *Cichorium intybus*; Ces, *Centaurea solstitialis*; Cl, *Curcuma longa*; Cm, *Centaurea maculosa*; Cr, *Ceratopteris richardii*; Cs, *Citrus sinensis*; Ct, *Carthamus tinctorius*; Ec, *Eleusine coracana*; Ee, *Euphorbia esula*; Eg, *Eucalyptus globulus*; Elg, *Elaeis guineensis*; Et, *Euphorbia tirucalli*; Fv, *Fragaria vesca*; Geh, *Gerbera hybrida*; Gm, *Glycine max*; Gh, *Gossypium hirsutum*, Gr, *Gossypium raimondii*; Ga, *Gossypium arboreum*; Ha, *Helianthus annuus*; Hp, *Helianthus petiolaris*; Hv, *Hordeum vulgare*; In, *Ipomoea nil*; Le, *Lycopersicon esculentum*; Lp, *Lycopersicon pennellii*; Ls, *Lactuca sativa*; Mc, *Mesembryanthemum crystallinum*; Md, *Malus* x *domestica*; Me, *Manihot esculenta*; Mg, *Mimulus guttatus*; Mt, *Medicago truncatula*; Nt, *Nicotiana tabacum*; Ob, *Ocimum basilicum*; Pa, *Prunus armeniaca*; Pg, *Picea glauca*; Ph, *Petunia* x *hybrida*; Pit, *Pinus taeda*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Pta, *Populus tremula* x *alba*; Pv, *Panicum virgatum*; Rc, *Ricinus communis*; Sea, *Senecio aethnensis*; Sec, *Senecio chrysanthemifolius*; Sb, *Sorghum bicolor*; Sh, *Saruma henryi*; So, *Saccharum officinarum*; St, *Solanum tuberosum*; Sc, *Solanum chacoense*; Ta, *Triticum aestivum*; To, *Taraxacum officinale*; Tk, *Taraxacum kok-saghyz*; Tr, *Tortula ruralis*; Vv, *Vitis vinifera*; Zo, *Zingiber officinale*; Zm, *Zea mays*. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.



Chez les arbres appartenant aux espèces de conifère (sous-classe des coniférospidées), les protéines PgLIM1a et b chez l'épinette blanche ou épicéa (*Picea glauca*) et PtLIM1a et b chez le pin à encens (*Pinus taeda*) sont très similaires dans leurs séquences. Pour ces deux espèces, les gènes *LIM1a* et *b* ont probablement subi un événement de duplication. Par ailleurs, il est difficile de les classer dans l'un des deux groupes  $\alpha$ LIM1 ou  $\beta$ LIM1. En effet, selon la méthode de construction des arbres basée sur le calcul d'une matrice de distance génétique et le regroupement des séquences par Neighbour-joining, les quatre protéines PgLIM1a et b, et PtLIM1a et b sont placées à l'extérieur de ces deux groupes et de fortes valeurs de bootstrap supportent cette séparation (Fig. 39). Il est important de souligner que, selon cette même méthode, la protéine CrLIM1 provenant de la fougère (*Ceratopteris richardii*) est également positionnée à l'extérieur de ces deux groupes et est proche des protéines LIM1 de conifère. Par contre, selon les méthodes de parcimonie et de maximum de vraisemblance pour la construction des arbres phylogénétiques, les protéines PgLIM1a et b, PtLIM1a et b, et CrLIM1 sont regroupées dans le nouveau groupe  $\beta$ LIM1, et dans ce cas, de fortes valeurs de bootstrap confortent aussi les noeuds (Annexe 4 et 5).

## Partie 1. Analyse phylogénétique



**Figure 40.** Arbres phylogénétiques du groupe  $\alpha$ LIM1 de la famille de protéines à domaine LIM chez les plantes. Les séquences d'acides aminés de 59 protéines appartenant au groupe  $\alpha$ LIM1 ont été alignées au moyen du logiciel ClustalW, puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leur appartenance aux groupes ou sous-groupes WLIM1, PLIM1, FLIM1, et  $\alpha$ LIM1 qui sont encadrés. Les différentes familles ou ordres de plante sont également encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

### 2.7.2. Le groupe $\alpha$ LIM1

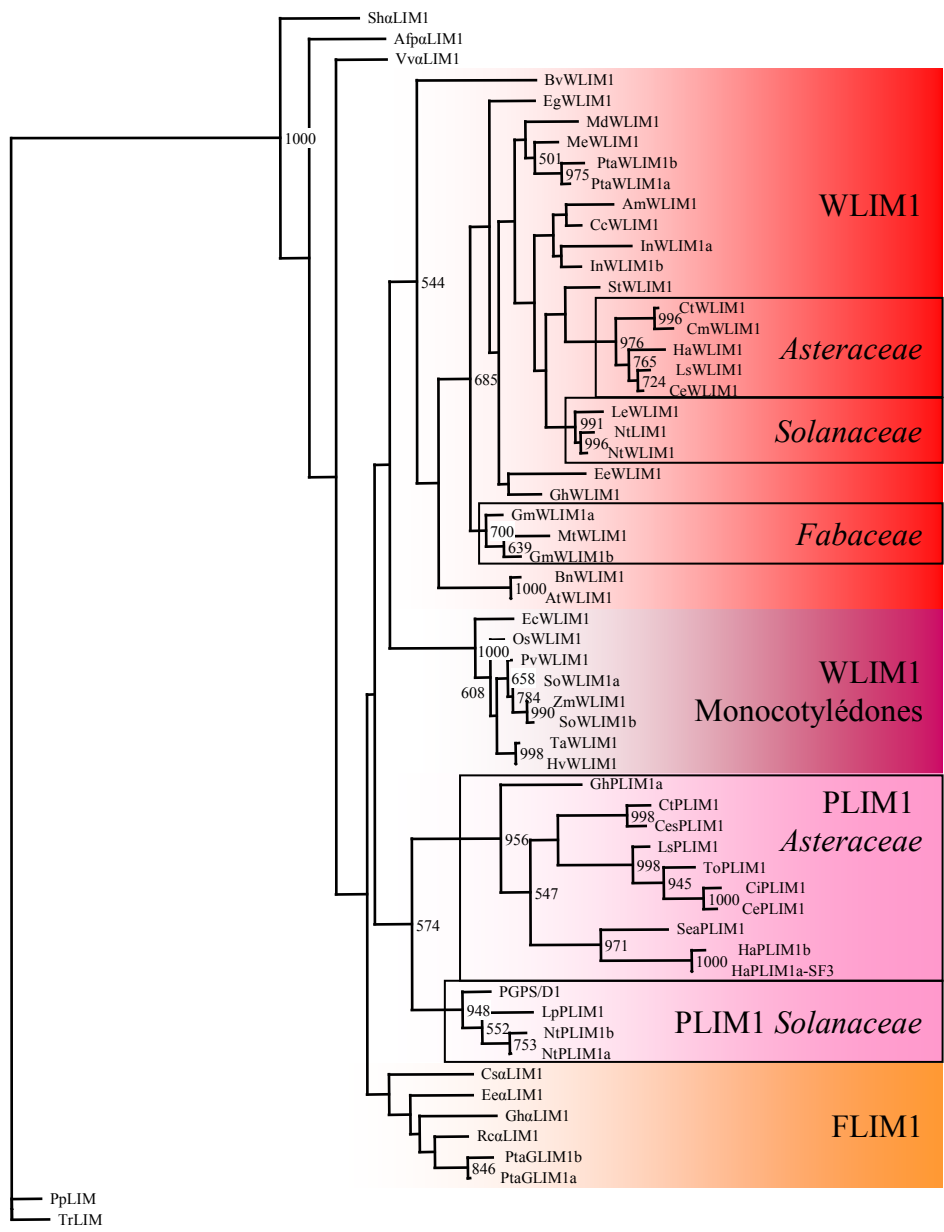
Les précédents sous-groupes identifiés WLIM1 et PLIM1 sont assemblés au sein du groupe  $\alpha$ LIM1 et globalement, quelque soit la méthode phylogénétique utilisée (distances génétiques, parcimonie et maximum de vraisemblance), de fortes valeurs de bootstrap confortent l'existence de ces deux sous-groupes (Fig. 40 et 41). De même, il paraît évident que le sous-groupe WLIM1 des monocotylédones forme un nouveau groupe monophylétique statistiquement significatif, et distinct du sous-groupe WLIM1 des dicotylédones. De plus, un quatrième sous-groupe appelé FLIM1 peut aussi être assigné au groupe  $\alpha$ LIM1. Toutefois les positions respectives des sous-groupes PLIM1, FLIM1 et WLIM1 des monocotylédones et des dicotylédones au sein du groupe  $\alpha$ LIM1 devraient être clarifiées. En effet, de par les faibles valeurs de bootstrap supportant les différents nœuds situés entre ces quatre sous-groupes, leurs origines phylogénétiques et leurs liens de parenté restent indéterminés. Les arbres Neighbour-joining générés à partir d'une matrice d'une distance calculée avec la méthode JTT ou à partir de la méthode de maximum de vraisemblance (PhyML) nous donnent un sous-groupe FLIM1 clairement séparé du sous-groupe PLIM1 (Fig. 40 et Fig. 41). Au contraire, quand la méthode de parcimonie est utilisée, les sous-groupes FLIM1 et WLIM1 des monocotylédones prennent place au sein du sous-groupe PLIM1, entre les protéines PLIM1 provenant des espèces de la famille des *Solanaceae* et celles provenant des espèces de la famille des *Asteraceae* (Fig. 41). De plus, cette dernière méthode favorise l'hypothèse d'un ancêtre commun pour les sous-groupes PLIM1 et WLIM1 des monocotylédones. Par ailleurs, certaines protéines à domaine LIM sont difficiles à classer au sein de ces sous-groupes du fait de leur appartenance à des ordres phylogénétiques (*Ranunculales*, *Piperales*, *Caryophyllales*, et *Vitales*) ancestraux aux sous-classes des rosidées et astérides. Ce sont les cas des protéines Afp $\alpha$ LIM1 de l'ancolie (*Aquilegia formosa*  $\times$  *A. pubescens*), ShaLIM1 de *Saruma henryi*, BvaLIM1 de betterave (*Beta vulgaris*), et VvaLIM1 de la vigne (*Vitis vinifera*).

Présentement, le sous-groupe FLIM1 n'est pas supporté par une valeur de bootstrap très élevée avec les méthodes de calcul de distances génétiques et de parcimonie, et les valeurs de bootstrap sont trop faibles avec la méthode de maximum de vraisemblance (Fig. 40 et Fig. 41A et 41B). De plus, il n'inclue qu'un faible nombre de protéines nommées PtGLIM1a et b, Gh $\alpha$ LIM1, EeaLIM1, et RcaLIM1. Comme nous le verrons plus tard, les noms PtGLIM1a et b incluent un préfixe G indiquant l'abondance des transcrits correspondants dans les fibres G de bois de tension.

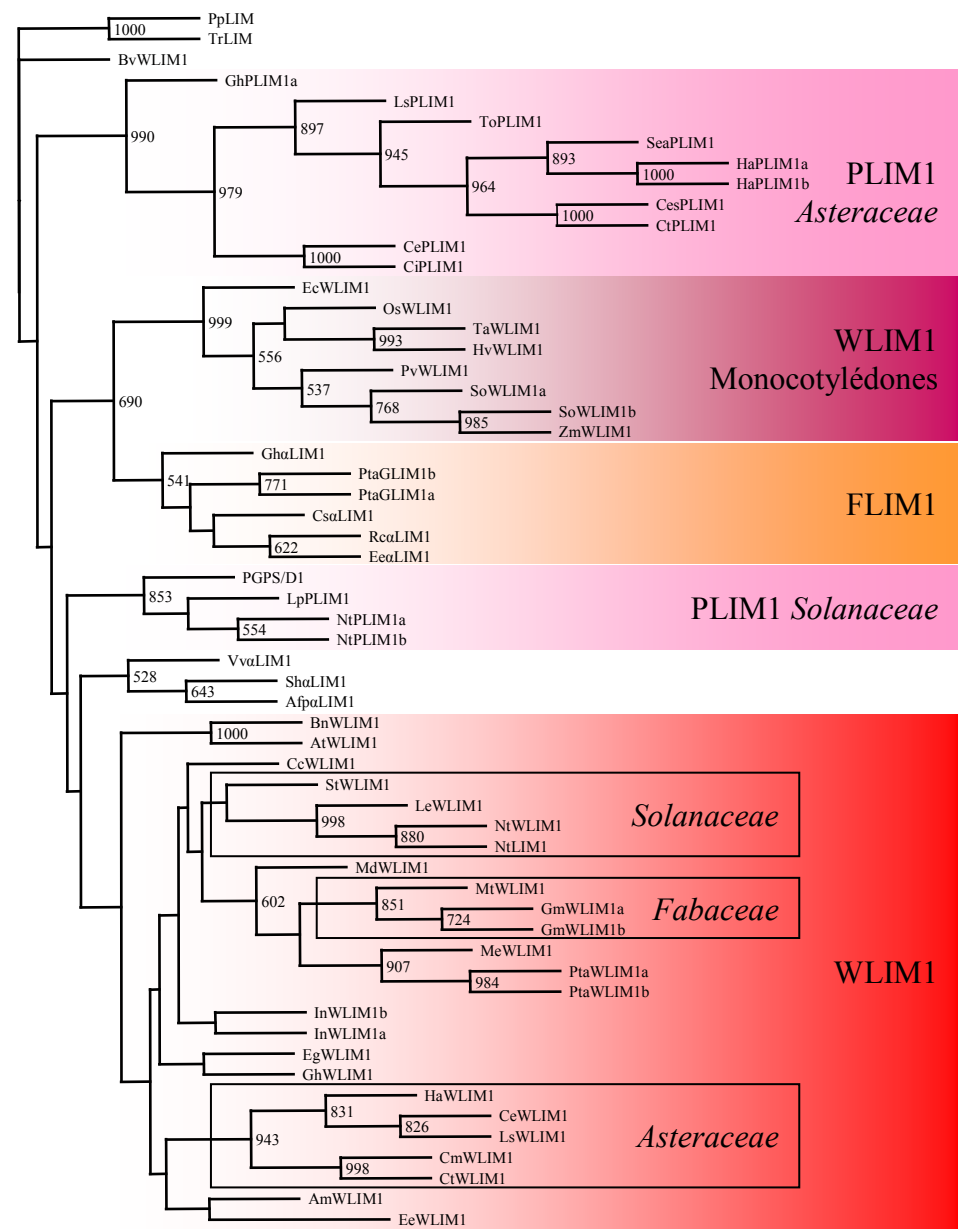
**A**

Partie 1. Analyse phylogénétique

**B**



0.1 substitution par site



100

## Partie 1. Analyse phylogénétique

**Figure 41.** Arbres phylogénétiques du groupe  $\alpha$ LIM1 de la famille de protéines à domaine LIM chez les plantes. Les séquences d'acides aminés de 59 protéines appartenant au groupe  $\alpha$ LIM1 ont été alignées au moyen du logiciel ClustalW, puis analysées soit avec la méthode de maximum de vraisemblance en utilisant les programmes Phylml et BIONJ (A), soit avec la méthode de parcimonie en utilisant le programme PHYLIP (B). Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. Pour l'arbre construit avec la méthode de maximum de vraisemblance, la longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leurs appartenances aux groupes ou sous-groupes WLIM1, PLIM1, FLIM1, ou  $\alpha$ LIM1 qui sont encadrés. Les différentes familles ou ordres de plante sont également encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

Les protéines PtGLIM1a et b de peuplier présentent 77 à 81 % d'identité en acides aminés avec GhaLIM1 du cotonnier (*Gossypium hirsutum*), 79 à 80 % avec la protéine EeaLIM1 issue de l'euphorbe (*Euphorbia esula*) et 87 à 89 % d'identité avec la protéine RcaLIM1 de ricin (*Ricinus communis*). Il n'est pas étonnant de constater de telles similitudes de séquences entre les protéines GLIM1 de peuplier et les protéines  $\alpha$ LIM1 d'euphorbes et de ricin. En effet, ces trois espèces appartiennent au même ordre des *Malpighiales*. Avec la méthode de parcimonie, la protéine CsaLIM1 issue de l'oranger (*Citrus sinensis*) est incluse dans le sous-groupe FLIM1, entre les protéines GLIM1 de peuplier et les protéines  $\alpha$ LIM1 de l'euphorbe et du ricin (Fig. 41B). Au contraire, avec les méthodes de calcul de distance génétique et de maximum de vraisemblance, cette protéine est proche du sous-groupe FLIM1 mais est positionnée à l'extérieur de ce groupe.

Il est intéressant de constater que le sous-groupe PLIM1 spécifique du pollen ne contient que des protéines à domaine LIM issues des familles de *Solanaceae* et *Asteraceae* de la sous-classe des astéridées (Fig. 40). Aucune protéine à domaine LIM provenant des espèces de monocotylédone, de conifères et de la sous-classe des rosidées, (incluant celles déduites des génomes séquencés d'*Arabidopsis*, de peuplier, et de riz) n'est classée au sein du sous-groupe PLIM1. De plus, les protéines PLIM1 des espèces de la famille des *Solanaceae* sont clairement séparées de celles appartenant aux espèces de la famille des *Asteraceae*. Cette séparation est plus nette lorsque l'on examine l'arbre phylogénétique construit avec la méthode de parcimonie : Les deux familles PLIM1 des *Asteraceae* et *Solanaceae* encadrent les sous-groupes FLIM1 et WLIM1 des monocotylédones. Les protéines PLIM1 issues de la famille des *Asteraceae* comprennent les protéines HaPLIM1a ou SF3 de tournesol (*Helianthus annuus*) (Baltz *et al.*, 1992a), GehPLIM1a de *Gerbera hybrida* (Laitinen *et al.*, 2005), LsPLIM1 de la laitue (*Lactuca sativa*), CePLIM1 de la chicorée endive (*Cichorium endivia*). Pour la famille des *Solanaceae*, on trouve les protéines NtPLIM1a et b de tabac



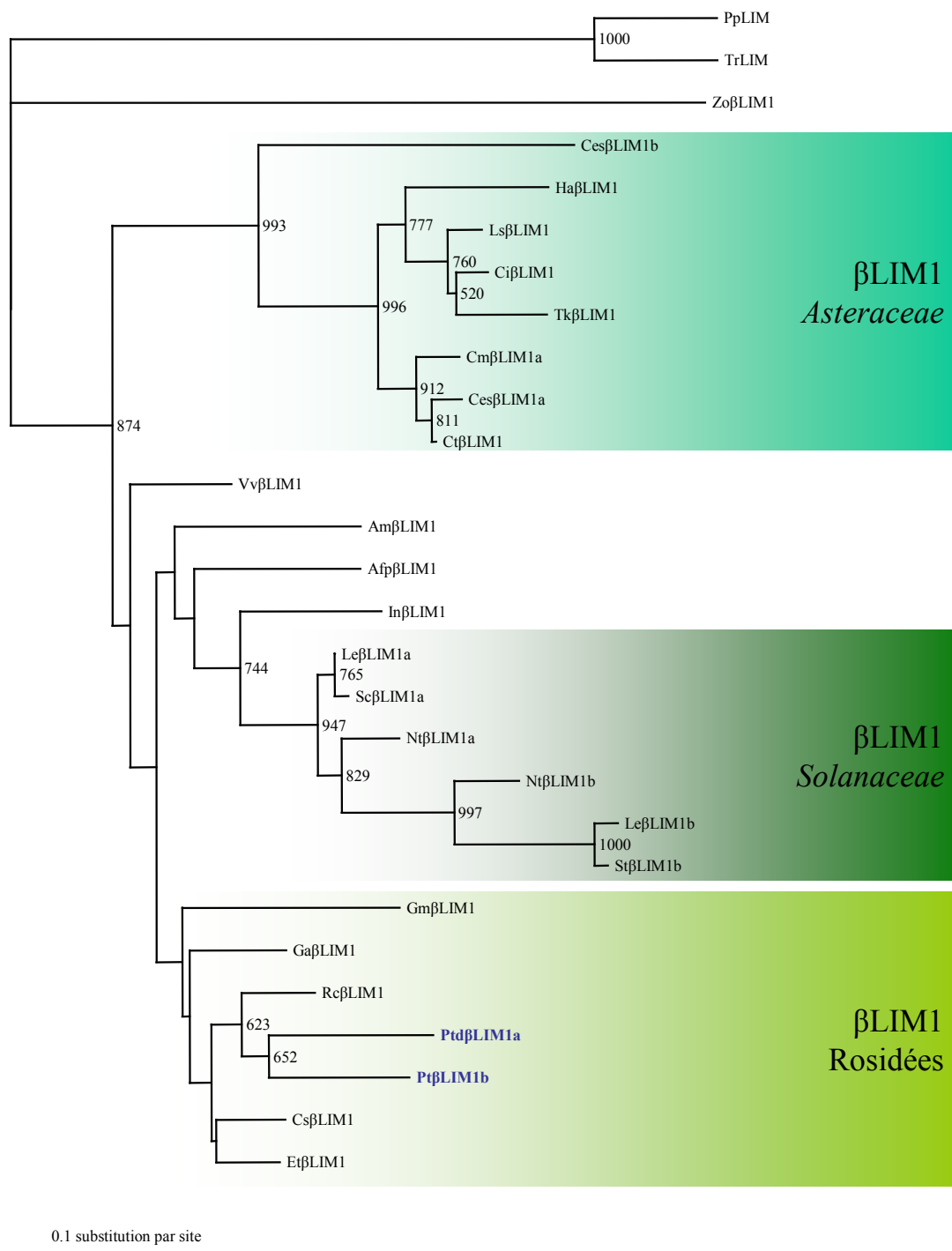
## Partie 1. Analyse phylogénétique

(*Nicotiana tabacum*) (Sweetman *et al.*, 2000), PGPS/D1 de pétunia (*Petunia × hybrida*) (Guyon *et al.*, 2000), et LpPLIM1 de tomate (*Lycopersicon pennellii*). Avec une grande longueur des branches, nous pouvons voir que les protéines du groupe PLIM1 ont fortement divergé entre elles.

Le sous-groupe WLIM1 des monocotylédones (classe des liliopsidées) contient des protéines fortement homologues entre elles. Ceci est illustré par un fort pourcentage d'identité entre elles, supérieur à 89 %, et une longueur de branche très courte entre les taxons. Ce pourcentage d'identité atteint même 99 % entre les protéines TaWLIM1 de blé (*Triticum aestivum*) et HvWLIM1 d'orge (*Hordeum vulgare*). Comme les deux transcrits *ZmWLIM1* identifiés sont probablement des allèles du même gène avec un seul acide aminé de différence entre les deux protéines, seul les protéines SoWLIM1a et b de canne à sucre (*Saccharum officinarum*), très proche entre elles avec une identité en acide aminé de 95 %, semblent être dupliquées au sein du sous-groupe WLIM1 des monocotylédones. On retrouve dans ce sous-groupe la protéine OsWLIM1 de riz.

Toutes les espèces de plantes issues du noyau des eudicotylédones (core eudicotylédones) sont représentées au sein du large sous-groupe WLIM1 des eudicotylédones (Fig. 40). Les protéines WLIM1 des eudicotylédones, par une plus grande longueur des branches les reliant, se sont plus diversifiées entre elles que les protéines WLIM1 des monocotylédones. Par contre, elles restent plus conservées entre elles que les protéines PLIM1. Au sein de ce sous-groupe, seul les protéines WLIM1 de peuplier, d'ipomée (*Ipomoea nil*), et de soja (*Glycine max*) semblent être dupliquées. Les deux gènes de tabac *NtWLIM1* et *NtLIM1* respectivement isolé par deux équipes différentes (Eliasson *et al.*, 2000 ; Kawaoka *et al.*, 2000) semblent être des allèles du même gène, car ils proviennent de deux variétés différentes de tabac Samsun et SR1 respectivement. La protéine NtLIM1 diffère de la protéine NtWLIM1, principalement par sept résidus acides aminés additionnels à l'extrémité C-terminale de la protéine. Chez l'eucalyptus (*Eucalyptus globulus*), le gène *EgLIM1* (AB208710) orthologue au gène *NtWLIM1* de tabac, et l'ADNc correspondant (AB208709) ont été soumis dans la base de donnée NCBI. Nous pouvons constater qu'au sein du sous-groupe WLIM1 de petits clades correspondant à différentes familles des sous-classes des rosidées ou astéridées, sont bien délimités. Ce sont la famille des *Fabaceae* avec les protéines GmWLIM1a et b et MtWLIM1 de soja et de luzerne (*Medicago truncatula*), et les familles des *Solanaceae* (tomate, tabac et pomme de terre) et des *Asteraceae* (tournesol, centaurée, laitue...).

## Partie 1. Analyse phylogénétique



**Figure 42.** Arbres phylogénétiques du groupe  $\beta$ LIM1 de la famille de protéines à domaine LIM chez les plantes. Les séquences d'acides aminés de 26 protéines appartenant au groupe  $\beta$ LIM1 ont été alignées au moyen du logiciel ClustalW, puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leur appartenance au groupe  $\beta$ LIM1. Les différentes familles ou sous-classe de plante sont encadrées. Les protéines à domaine LIM de peuplier sont marquées en bleu gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

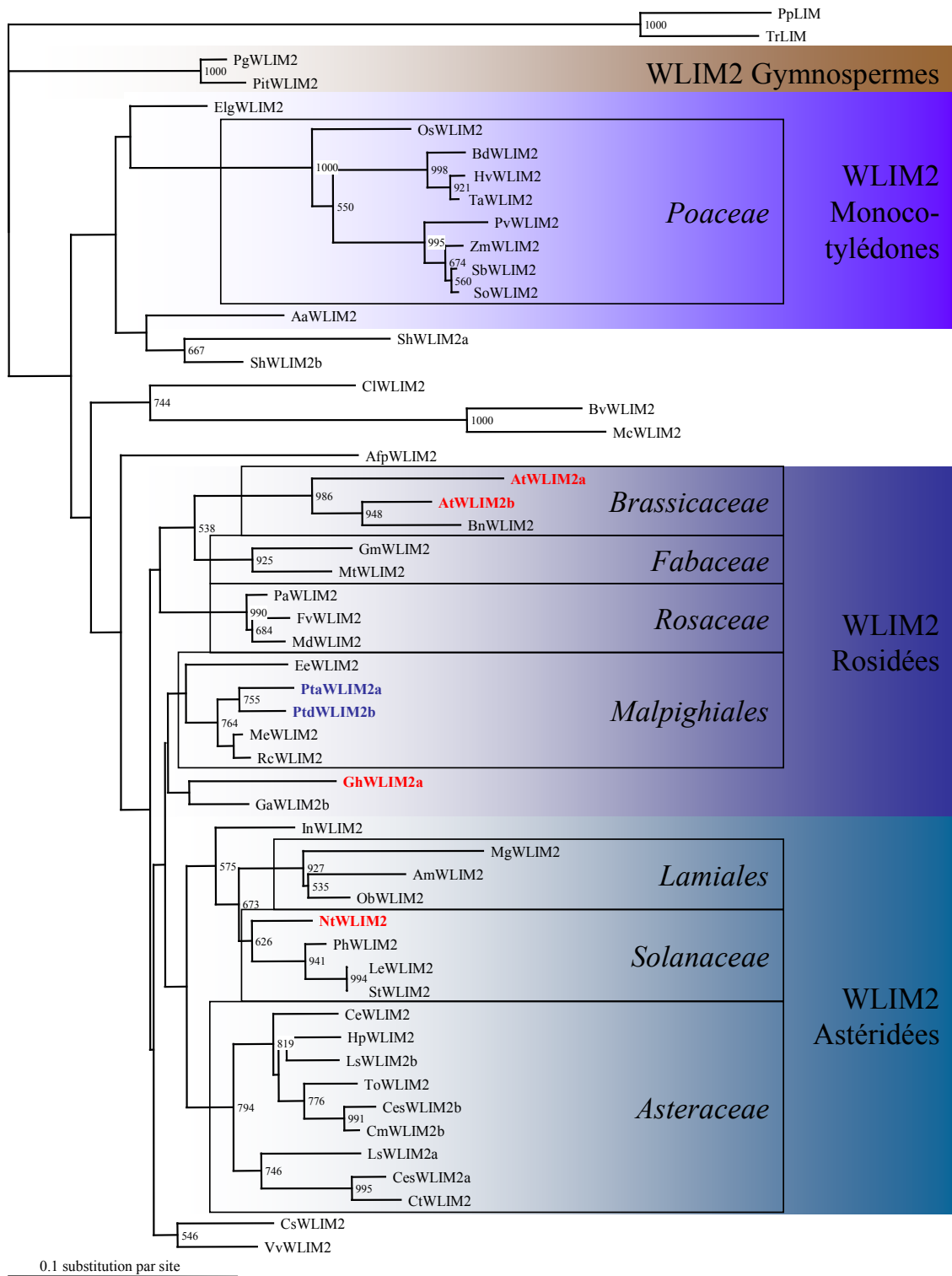
### 2.7.3. Le nouveau groupe $\beta$ LIM1

Le nouveau groupe  $\beta$ LIM1 identifié contient peu de protéines à domaine LIM comparé aux autres groupes LIM. Elles sont issues d'espèces appartenant aux sous-classes des astéridées et rosidées, incluant les protéines Pt $\beta$ LIM1a et b de peuplier (Fig. 42). De manière intéressante, il n'y a pas de protéines LIM provenant d'*Arabidopsis* ni de riz dans ce groupe. De la même manière aucune protéine à domaine LIM issues d'espèces de la famille des *Brassicaceae* (ou crucifère) et d'espèces de la famille des *Poaceae* (blé, maïs...) n'a été identifiée au sein de ce groupe  $\beta$ LIM1. La protéine Zo $\beta$ LIM1 issue du gingembre (*Zingiber officinale*) de la famille des *Zingiberaceae* est la seule protéine identifiée dans le groupe  $\beta$ LIM1 appartenant à la classe des liliopsidées (ou monocotylédones). Par ailleurs, quelque soit la méthode de construction d'arbres phylogénétiques utilisée (distance génétique, parcimonie, ou maximum de vraisemblance), les protéines issues des espèces des familles de *Solanaceae* et *Asteraceae* forment deux petits sous-groupes monophylétiques séparés à l'intérieur du groupe  $\beta$ LIM1. Parmi les protéines issues d'espèces de la sous-classe des astéridées, seul les protéines issues d'espèces de la famille des *Solanaceae* (tabac, tomate et pomme de terre) semblent être dupliquées. Les deux protéines Ces $\beta$ LIM1a et b, issues de la centaurée du solstice (*Centaurea solstitialis*), semblent en effet être trop éloignées phylogénétiquement pour être le résultat d'un événement de duplication récent.

### 2.7.4. Le groupe $\gamma$ LIM2

Une forte valeur de bootstrap marque la séparation phylogénétique des groupes  $\gamma$ LIM2 et  $\delta$ LIM2 (Fig. 39). Le groupe  $\gamma$ LIM2 est un grand groupe comprenant beaucoup de protéines et contient les deux sous-groupes WLIM2 de monocotylédone et eudicotylédone. Ces deux sous-groupes sont eux aussi nettement séparés l'un de l'autre par une forte valeur de bootstrap. Au sein de chaque sous-groupe WLIM2 les longueurs des branches, très courtes, indiquent une forte conservation des protéines WLIM2. Les longueurs des branches entre les protéines WLIM2 de monocotylédones ou de eudicotylédones sont globalement plus courtes que celles observées entre les protéines WLIM1 des eudicotylédones (Fig. 43). Les protéines WLIM2 des eudicotylédones sont donc plus conservées entre elles que les protéines WLIM1 des eudicotylédones, et les protéines WLIM2 des monocotylédones semblent autant conservées entre elles que les protéines WLIM1 des monocotylédones.

## Partie 1. Analyse phylogénétique



**Figure 43.** Arbres phylogénétiques du groupe  $\gamma$ LIM2 de la famille de protéines à domaine LIM chez les plantes. Les séquences d'acides aminés de 52 protéines appartenant au groupe  $\gamma$ LIM2 (ou sous-groupe WLIM2) ont été alignées au moyen du logiciel ClustalW, puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leur appartenance aux sous-groupes WLIM2 qui sont encadrés. Les différentes familles ou ordres de plante sont également encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

## Partie 1. Analyse phylogénétique

Les protéines ElgWLIM2 issue de palmier à huile (*Elaeis guineensis*) de la famille des *Arecaceae* et AcWLIM2 issue de l'acore commun (*Acorus americanus*) de la famille des *Acoraceae* sont des protéines provenant de familles de monocotylédones (liliopsidées) plus anciennes que la famille des *Poaceae*. Effectivement, ces deux protéines sont relativement éloignées des protéines WLIM2 appartenant à la famille des *Poaceae*, mais elles restent phylogénétiquement bien distinctes des protéines du sous-groupe WLIM2 des eudicotylédones (Fig. 43). Étonnamment, la protéine CIWLIM2 issue du curcuma (*Curcuma longa*) est clairement séparée des protéines WLIM2 de monocotylédones, alors que le curcuma de la famille des *Zingiberaceae* appartient aussi à la classe des liliopsidées, comme les espèces de la famille des *Poaceae*. Avec les méthodes de calcul de distances génétiques et de maximum de vraisemblance, CIWLIM2 est proche des protéines BvWLIM2 et McWLIM2 provenant des espèces *Beta vulgaris* et *Mesembryanthemum crystallinum* de l'ordre des *Caryophyllales*. Par contre, avec la méthode de parcimonie, la protéine CIWLIM2 est franchement séparée des protéines BvWLIM2 et McWLIM2 et est plus proche de la protéine ElgWLIM2 issue du palmier. On peut constater que les conifères *Picea glauca* et *Pinus taeda* possèdent aussi des protéines WLIM2 appartenant au groupe  $\gamma$ LIM2, mais contrairement aux gènes dupliqués *PitLIM1a* et *b* et *PgLIM1a* et *b*, des gènes dupliqués *WLIM2* pour ces deux espèces n'ont pas été trouvés. De par des valeurs de bootstrap significativement fortes, au sein du sous-groupe WLIM2 des eudicotylédones, les protéines WLIM2 issues des espèces de la sous-classe des astéridées sont nettement différentes de celles issues des espèces de la sous-classe des rosidées. De même, au sein de la sous-classe des astéridées, les protéines issues des espèces des ordres des *Solanales* (ipomée, tomate, tabac...) et *Lamiales*, comme les protéines AmWLIM2 et ObWLIM2 de muflier (*Antirrhinum majus*), et de basilic (*Ocimum basilicum*), sont distinctes de celles issues de l'ordre des *Asterales* (tournesol, pissenlit, laitue...). Pareillement, dans la sous-classe des rosidées, les protéines issues d'espèces de l'ordre des *Malpighiales*, tels que PtWLIM2a et b de peuplier, RcWLIM2 de ricin, EeWLIM2 de l'euphorbe, et MeWLIM2 du manioc (*Manihot esculenta*), ainsi que celles issues d'espèces de la famille des *Rosaceae* (ordre des *Rosales*), tels que FvWLIM2 du fraisier (*Fragaria vesca*), MdWLIM2 de pommier (*Malus × domestica*) et PaWLIM2 de l'abricotier (*Prunus armeniaca*), forment des petits sous-groupes au sein du sous-groupe WLIM2. On constate aussi de petits sous-groupes de protéines issues des espèces des familles de *Brassicaceae* et *Fabaceae*. Dans ce sous-groupe WLIM2 de la sous classe des rosidées, la protéine du cotonnier initialement nommée GhLIM1 (Luo *et al.*, 2003), a été renommée GhWLIM2a. Pour certaines plantes eudicotylédones comme *Arabidopsis*, le peuplier, le cotonnier, et la

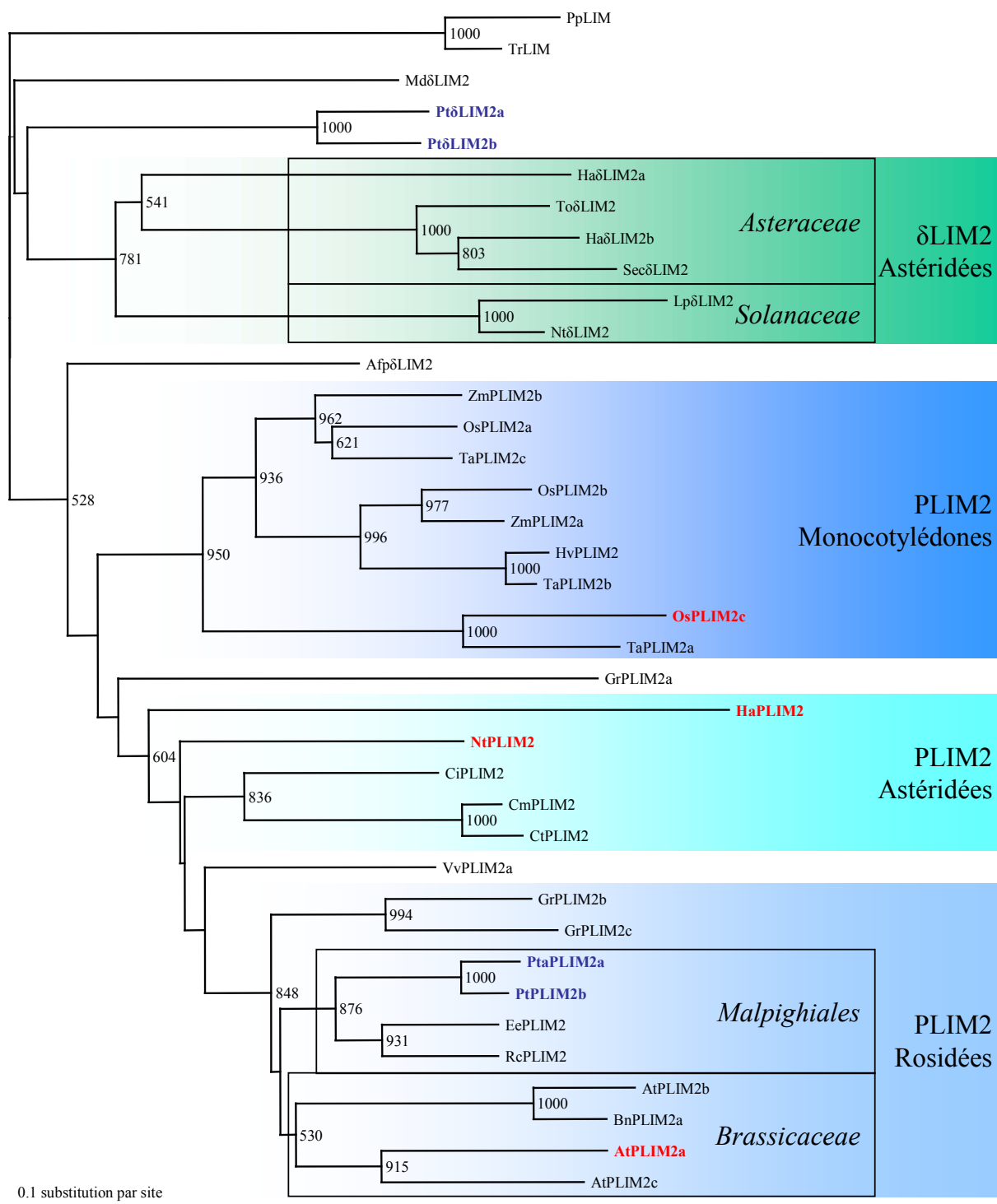
laitue, les gènes *WLIM2* sont dupliqués, alors que cela n'est jamais le cas pour les plantes monocotylédones.

### 2.7.5. Le groupe $\delta$ LIM2

Par le nombre important de séquences de protéines à domaine LIM identifiées, le groupe  $\delta$ LIM2 est considérablement agrandi. Quelque soit la méthode de construction de l'arbre utilisé, le groupe  $\delta$ LIM2 est principalement divisé en trois sous-groupes monophylétiques : les sous-groupes PLIM2 des eudicotylédones, PLIM2 des monocotylédones, et  $\delta$ LIM2 des astéridées qui contient notamment les séquences des nouvelles protéines identifiées chez le tournesol et le tabac (Fig. 44). Plusieurs événements de duplication se sont produits au sein des gènes  $\delta$ LIM2, montrant une importante diversification au sein de ce groupe durant l'évolution. De par la longueur très longue des branches, les protéines  $\delta$ LIM2 sont en moyenne, plus divergentes entre elles que les protéines issues des autres groupes LIM de plante. Les protéines PLIM2 des eudicotylédones peuvent facilement être divisées en deux sous-groupes monophylétiques, l'un comprenant des protéines PLIM2 issues d'espèces de la sous-classe des astéridées et l'autre contenant des protéines issues d'espèces de la sous-classe des rosidées. Dans cette dernière sous-classe, on trouve des protéines PLIM2 dupliquées chez le peuplier, le coton et *Arabidopsis* qui en contient trois. Par ailleurs, deux petits groupes comprenant des protéines issues d'une part d'espèces de l'ordre des *Malpighiales* (peuplier, euphorbe et ricin) et d'autre part d'espèces de la famille des *Brassicaceae* (*Arabidopsis* et colza) sont bien distinguables.

Les protéines PLIM2 des monocotylédones sont phylogénétiquement séparées des protéines PLIM2 des eudicotylédones, et une forte valeur de bootstrap supporte cette séparation (Fig. 44). Chez les monocotylédones, les protéines PLIM2 sont généralement trouvées en triplicat, résultant au moins de deux événements de duplication, et forment ainsi trois petits groupes au sein du sous-groupe PLIM2 des monocotylédones. Certaines protéines déduites d'orge et de maïs, bien qu'appartenant à ce sous-groupe PLIM2, n'ont pas pu être intégrées dans l'arbre phylogénétique car elles étaient tronquées. Contrairement aux protéines WLIM2 de monocotylédones très conservées, la longueur des branches indique que les protéines PLIM2 de monocotylédones sont beaucoup plus divergentes entre elles.

## Partie 1. Analyse phylogénétique



**Figure 44.** Arbres phylogénétiques du groupe  $\delta$ LIM2 de la famille de protéines à domaine LIM chez les plantes. Les séquences d'acides aminés de 36 protéines appartenant au groupe  $\delta$ LIM2 ont été alignées au moyen du logiciel ClustalW, puis analysées par la méthode de Neighbour-joining avec des matrices de distances génétiques calculées en utilisant le modèle JTT de changement en acides aminés. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leur appartenance aux groupes ou sous-groupes  $\delta$ LIM2 et PLIM2 qui sont encadrés. Les différentes familles ou ordres de plante sont également encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

## Partie 1. Analyse phylogénétique

Les gènes  $\delta LIM2$  et  $PLIM2$  des espèces appartenant à la sous-classe des astéridées semblent aussi dupliqués au sein du groupe  $\delta LIM2$  (Fig. 44). Mais, contrairement aux gènes  $PLIM2$  d'*Arabidopsis* relativement proches entre eux et trouvés en triplicat au sein du sous-groupe  $PLIM2$  des eudicotylédones, les nouvelles protéines  $\delta LIM2$  de tabac et de tournesol sont plus éloignées phylogénétiquement par rapport aux protéines Ha $PLIM2$  et Nt $PLIM2$  précédemment identifiées (Eliasson *et al.*, 2000). Ces protéines Nt $\delta LIM2$ , Ha $\delta LIM2a$  et Ha $\delta LIM2b$  sont donc groupées dans un nouveau sous-groupe monophylétique appelé  $\delta LIM2$  des astéridées. Ce nouveau sous-groupe reste très proche des sous-groupes  $PLIM2$  des eudicotylédones et monocotylédones. Les séquences protéiques Pt $\delta LIM2a$  et b déduites à partir du génome de *Populus trichocarpa* sont phylogénétiquement proches du sous-groupe  $\delta LIM2$  des astéridées, et relativement éloignées des deux sous-groupes  $PLIM2$ . Enfin les deux protéines Afp $\delta LIM2$  et Md $\delta LIM2$  de l'ancolie et du pommier sont difficiles à positionner dans l'un de ces trois sous-groupes.

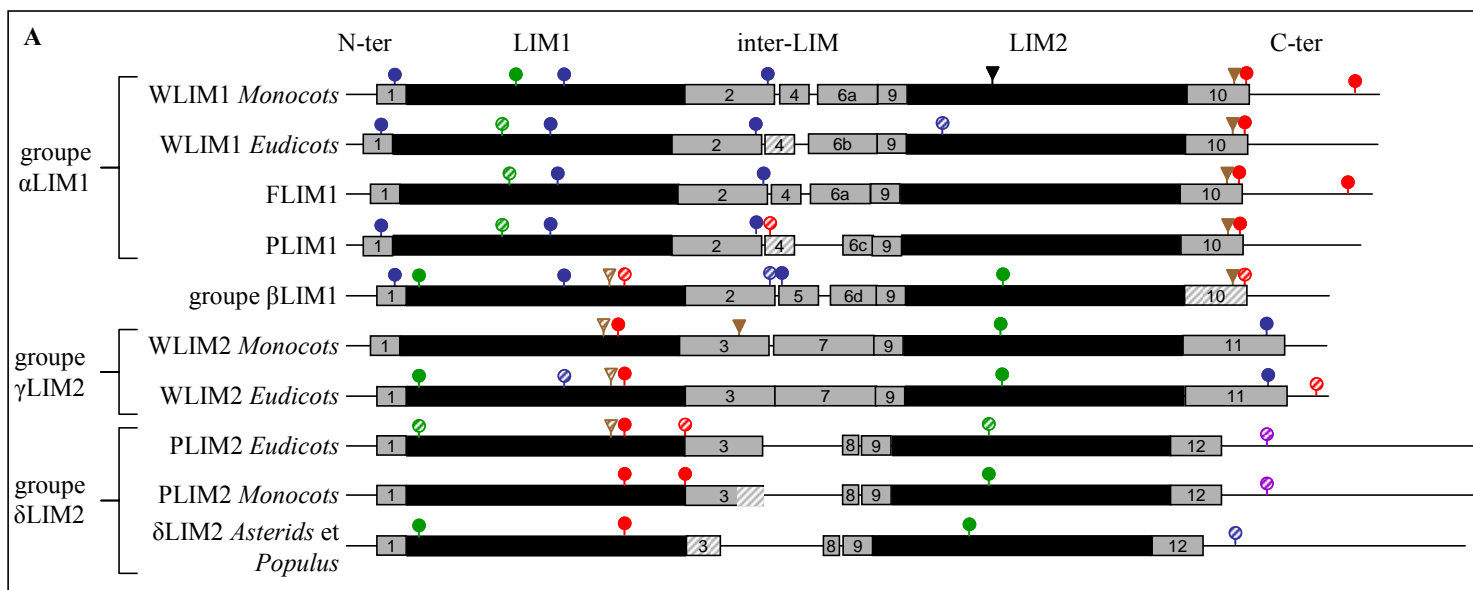
### 2.8. Analyse de séquences et de la structure des protéines à domaine LIM de plante

#### 2.8.1. Comparaison des séquences des domaines LIM entre les protéines CRP animales et les protéines à domaine LIM de plante

Les protéines à domaine LIM de plante présentent beaucoup de caractéristiques communes avec les protéines riches en cystéine animales (protéines CRP), mais aussi plusieurs spécificités. Par exemple, les deux régions riches en glycine (GRR) suivant chaque domaine LIM chez les protéines CRP ne sont pas retrouvées dans les protéines à domaine LIM de plante (Eliasson *et al.*, 2000). Toutefois, pour toutes les protéines LIM de plante, un résidu glycine est strictement conservé neuf acides aminés après le dernier ligand au zinc pour chaque domaine LIM (Fig. 45B et Annexe 6). De même, en amont de chaque domaine LIM, une autre glycine est aussi fortement conservée, cinq acides aminés avant le premier ligand au zinc. D'autre part, le signal potentiel d'adressage nucléaire (KKYGPK) présent dans les protéines de la famille CRP chez les vertébrés est aussi manquant dans les protéines LIM de plante. Enfin, le domaine C-terminal des protéines CRP animales est très court, contrairement à celui des protéines à domaine LIM de plante.



## Partie 1. Analyse phylogénétique



## B

### Domaine LIM1

$C-X_2-C-X\{K/T-V-(Y/H)\}X_2-(D/E)-X-L-X_2-(D/E)-X_3-(Y/F)-H-(K/R)-X-C-F-(R/K)-C-X-H-C-X_3-L-X_4-(Y/F)-X-S-X-(E/D)-G-V-(L/P)-(Y/F)-C-(K/R)-X-H$

### Domaine LIM2

$C-X_2-C-X\{K/X\}-T-(V/A)-Y\}P-X-E-(K/R)-X_7-(Y/F)-H-(K/R)-X-C-F-(K/R)-C-X-(H/X)-G_2-C-X-(I/L)-(S/T)-X-S-(N/S)-Y-X_4-(G/H)-X-L-(Y/F)-C-(K/R)-(H/X)-H$

### Autres motifs conservés dans les protéines à domaine LIM de plante

(1)  $F-X_{0-1}-G-T-X_2-K$

(2)  $(F/Y)DQLFK(R/X)T\text{G}SL(D/E)KSF(E/D)G(T/X)P(K/R)$

(3)  $F(E/D)QL(F/Y)KE(T/S)\text{G}(N/S)(F/Y)(N/S)K((N/S)F(Q/X)(S/X)$

(4)  $K(P/X)(E/X)(R/K)$

(5)  $TVR_{X_{0-1}}(E/D)RS(A/X)D$

(6a)  $(G/D)NENAXKVS(S/N)$

(6b)  $(D/E)(G/X)E(K/R)PX3KVS(S/N)$

(6c)  $K(V/X)S(S/N)$

(6d)  $(S/T)(N/X)(S/T)KVSS$

(7)  $PAK_{X_{2-3}}(E/D)K(L/X)(T/X)P(E/M)LT(R/K)SPSKAA$

(8)  $(R/K)XP(S/N)K$

(9)  $F-X-G-T-(Q/R)-(D/E)-K$

(10)  $H(I/X)QL(I/F)K(E/X)K\text{G}N(L/F/Y)S(Q/K)LE$

(11)  $F(S/A)QLFKEK\text{G}SYNHLIK(C/S)ASXKR$

(12)  $F(A/X)QLF(M/K)(E/X)(K/X)\text{G}(S/X)(Y/X)$

## Partie 1. Analyse phylogénétique

**Figure 45.** Identification de motifs conservés dans les protéines LIM de plante et définition des séquences consensus correspondantes.

(A) Représentation schématique des différents motifs conservés dans les séquences des protéines à domaine LIM de plante. Les séquences des protéines LIM ont été regroupées en fonction de leurs appartenances aux groupes ou sous-groupes LIM, de manière à assembler les protéines orthologues entre elles. Les abréviations *Monocots* et *Eudicots* correspondent respectivement aux espèces monocotylédones et eudicotylédones. Les domaines LIM1 et LIM2 sont représentés par des rectangles noirs. Les motifs conservés entourant les domaines LIM et identifiés avec les programmes ClustalW et MEME sont représentés par des rectangles gris. Les sites putatifs de glycosylation sur les asparagines (ASN) sont repérés par des triangles bruns. Les sites putatifs de phosphorylation sur les tyrosines (Tyr), ainsi que les sites putatifs de phosphorylation par les caséines kinase (CKII), les protéines kinase C (PKC), et les protéines kinase dépendantes du cAMP et du cGMP (cAMP), sont respectivement indiqués par des ronds verts, oranges, bleus et violets. Les motifs de glycosylation ou de phosphorylation non conservés sont hachurés.

(B) Séquences consensus des domaines LIM1 et LIM2 de plante et des quinze motifs identifiés avec le programme MEME. Les cystéines et les histidines liant les atomes de zinc sont marquées en gras. Les première et dernière glycines conservées avant et après chaque domaine LIM sont soulignées. Le symbole X représente n'importe quel acide aminé.



Les domaines LIM sont très conservés entre les protéines CRP animale et les protéines LIM de plante. Chez les plantes, le domaine LIM contient deux doigts de zinc répétés en tandem avec la structure caractéristique suivante [C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-H] (Eliasson *et al.*, 2000) (Fig. 45). Contrairement aux protéines de la famille CRP animale, chez les plantes, le dernier ligand au zinc est remplacé par une histidine pour chaque domaine LIM. De plus, le second domaine LIM chez les plantes est atypique avec la cystéine du second ligand du deuxième doigt de zinc remplacé par une glycine conservée. Par contre, pour toutes les protéines LIM de plante on trouve, pour le second domaine LIM, une cystéine strictement conservée pouvant peut-être permettre ainsi la coordination au zinc. Elle est positionnée cinq acides aminés après le premier ligand au zinc du deuxième doigt de zinc (Fig. 45). De façon surprenante, la première cystéine du premier domaine LIM est manquante (remplacée par une arginine) seulement pour la protéine dupliquée SoWLIM1b de canne à sucre, et chez le maïs pour la seconde version allélique de la protéine ZmWLIM1 représentée par l'ADNc AY112454 (résultats non montrés). La protéine codée par cet ADNc diffère de celle utilisée dans l'analyse phylogénétique uniquement par l'absence de cette cystéine. Chez les deux familles CRP animale et LIM de plante, avec l'exception des protéines du sous-groupe PLIM2 des monocotylédones, un motif K[T/A]VY fortement conservé est trouvé dans le premier domaine LIM, proche de la deuxième cystéine du premier doigt de zinc (Fig. 45B). Ce motif est aussi trouvé, mais moins conservé, à la même position au niveau du second domaine LIM. Chez les plantes, mais de façon moins évidente chez les animaux, les deux doigts de zinc de chaque domaine LIM sont liés par deux acides aminés très conservés F et [R/K]. En général, les acides aminés entourant les deux domaines LIM sont aussi très conservés entre les protéines LIM de chaque sous-groupe, avec l'exception notable des

protéines du sous-groupe  $\beta$ LIM2 (Fig. 45A et 45B). Enfin, nous pouvons constater que le domaine LIM1 est en général plus conservé que le domaine LIM2. Toutefois, pour les protéines WLIM2, les domaines LIM sont tous deux fortement conservés.

### 2.8.2. Spécificités structurales des protéines à domaines LIM au sein de chaque sous-groupe phylogénétique et identification de sites putatifs de phosphorylation et glycosylation

Les protéines LIM peuvent avoir de multiples partenaires dans la cellule, comme le cytosquelette d'actine, l'ADN ou d'autres protéines potentielles. Bien que leurs fonctions précises soient inconnues, il a été proposé que les protéines LIM de plante puissent servir comme navette entre le cytoplasme et le noyau. Pour assurer cette double fonction il est très probable qu'elles soient régulées par différentes modifications post-traductionnelles telles que des phosphorylations et glycosylations. Dans ce cadre, j'ai cherché des sites putatifs de phosphorylation et de glycosylation dans les séquences des protéines à domaine LIM de plante, en utilisant la base de donnée Prosite. Grâce au grand nombre de séquences collectées, nous sommes capable de réaliser des comparaisons de séquences orthologues au sein des sous-groupes LIM pour trouver des motifs conservés avec le programme MEME.

Globalement, la plupart des différences observées entre ces protéines à domaine LIM sont concentrées à la fois dans les domaines inter-LIM et C-terminal (Fig. 45). Au sein du groupe  $\alpha$ LIM1, les protéines du sous-groupe PLIM1 sont plus divergentes entre elles que les protéines des sous-groupes FLIM1, WLIM1. Le domaine C-terminal des protéines du sous-groupe PLIM1 est plus différent que celui des protéines WLIM1 et contient des résidus proline additionnels (Annexe 6). En remarque, la répétition du pentapeptide « ETQNA » situé dans le domaine C-terminal des protéines HaPLIM1a et b n'est pas conservé pour les autres protéines du sous-groupe PLIM1 de la famille des *Asteraceae*. En revanche, un autre motif répété « PEAA » est trouvé dans le domaine C-terminal de la protéine CtPLIM1 de *Carthamus tinctorius*. Les protéines du sous-groupe FLIM1 ressemblent plus aux protéines du sous-groupe WLIM1 qu'aux protéines du sous-groupe PLIM1, particulièrement au niveau des domaines LIM. Le domaine inter-LIM des protéines FLIM1 est similaire à celui des protéines du sous-groupe WLIM1 des monocotylédones, alors que leur domaine C-terminal diffère nettement de celui des autres sous-groupes WLIM1 et PLIM1. Au niveau des domaines LIM, les protéines du groupe  $\beta$ LIM1 portent des caractéristiques structurales similaires à la fois aux protéines du groupe  $\alpha$ LIM1 et aux protéines du groupe  $\gamma$ LIM2, mais les domaines inter-LIM

et C-terminal sont spécifiques de ce groupe. Excepté pour les protéines WLIM2 issues d'espèces de la classe des conifères et de l'ordre des *Caryophyllales*, les protéines du sous-groupe WLIM2 ont globalement des séquences fortement conservées entre elles, même au niveau du domaine inter-LIM. Les différences qui distinguent les protéines WLIM2 de monocotylédones et de eudicotylédones sont principalement localisées dans les domaines LIM1, inter-LIM et C-terminal. Le groupe  $\delta$ LIM2 est le groupe le plus hétérogène même au niveau des domaines LIM. Les protéines du groupe  $\delta$ LIM2 sont caractérisées par un domaine C-terminal très long et fortement variable dans sa composition en acides aminés. Ce domaine C-terminal est majoritairement composé d'acides aminés acides, particulièrement des résidus d'acides glutamiques (Annexe 6). Pour ces protéines  $\delta$ LIM2, le point isoélectrique (pI) de ce domaine C-terminal varie entre 4 et 6. Contrairement aux protéines  $\alpha$ LIM1 et  $\delta$ LIM2, les protéines du sous-groupe WLIM2 ont un domaine C-terminal basique qui ne contient pas d'acides aminés polaires, excepté un acide aminé acide conservé à la fin des protéines. Les domaines C-terminal des protéines du groupe  $\beta$ LIM1 ont des pI variables. Le pI est plutôt basique pour les protéines dupliquées issues des espèces de la famille des *Solanaceae*, et acide ou neutre pour les protéines issues d'espèce d'*Asteraceae* (résultats non montrés).

Différents types de motifs comme des sites putatifs de glycosylation sur les asparagines, des sites putatifs de phosphorylation sur les tyrosines (Tyr), ou des sites putatifs de phosphorylation par les caséines kinase II (CKII), et par les protéines kinase C (PKC) sont localisés à l'intérieur et à l'extérieur des domaines LIM (Fig. 45). Par exemple, dans les protéines des groupes  $\alpha$ LIM1 et  $\beta$ LIM1, et seulement pour ces protéines, un site putatif de phosphorylation PKC est trouvé juste avant le premier ligand au zinc du domaine LIM1. Toutefois, les protéines du sous-groupe FLIM1 (appartenant au groupe  $\alpha$ LIM1) ne possèdent pas ce site putatif de phosphorylation. Finalement, un autre site putatif de phosphorylation PKC très conservé est exclusivement rencontré dans le domaine C-terminal des protéines WLIM2 issues des angiospermes et aussi des gymnospermes. Ce site de phosphorylation PKC est suivi d'une région riche en alanine qui est particulièrement longue pour les protéines WLIM2 issues d'espèces des familles d'*Asteraceae* et de *Brassicaceae*. Après la région riche en alanine du domaine C-terminal des protéines WLIM2 des eudicotylédones, un site de phosphorylation CKII partiellement conservé est trouvé. De manière intéressante, ce site a été perdu pour les protéines dupliquées AtWLIM2a, PtWLIM2a et LsWLIM2a et n'est pas présent dans les protéines WLIM2 issues d'espèces de monocotylédones.

### 3 Discussion

#### 3.1. Les SNPs observés dans les gènes LIM de peuplier peuvent avoir une influence sur la fonction des protéines

Les polymorphismes simples de nucléotides (SNPs) correspondent à des mutations naturelles apparues entre les différentes espèces ou hybrides de peuplier. Elles reflètent sûrement l'importante diversité des phénotypes morphologiques ou écophysiologiques observés chez le genre *Populus* (Eckenwalder, 1996). En effet, le nombre d'espèces et sous-espèces de peuplier est important, sa répartition géographique est large (Europe, Asie, Amérique...) et son habitat est très diversifié. Chez *Populus tremula*, un taux élevé de polymorphisme nucléotidique au niveau des régions codantes et non codantes a été identifié pour cinq gènes au sein des populations européennes et avec *P. trichocarpa* (Ingvarsson, 2005). Le niveau de polymorphisme chez *P. tremula* est deux à dix fois plus élevé que celui observé chez d'autres espèces ligneuses et pérennes comme le pin. Par ailleurs, il est important de préciser que ces SNPs ne semblent pas résulter d'erreurs de la Taq Polymérase. En effet, le taux moyen d'erreurs de cette enzyme ( $8 \pm 3,6 \times 10^{-6}$ ) (Cline *et al.*, 1996), n'explique pas le nombre de mutations en commun entre les différentes ESTs. J'ai identifié des SNPs pour les gènes PtaWLIM1a et PtaGLIM1a au sein des populations d'ESTs séquencées à partir de l'hybride *Populus tremula* × *P. alba*. Toutefois, des analyses génétiques plus poussées doivent être réalisées pour déterminer l'origine parentale des formes alléliques chez l'hybride *Populus tremula* × *P. alba*. Pour les gènes *PtaGLIM1b* et *PtaWLIM1b*, je n'ai pas identifié d'autre allèle au sein des populations d'ESTs de ces deux consensus. Ces deux gènes sont probablement mono-allélique dans ce fond génétique (*P. tremula* × *P. alba*) ou alors, il n'y a pas suffisamment d'EST pour définir un autre allèle. Par contre, des SNPs ont été détectés au sein des séquences codantes, entre les six ADNc *PtaLIM* de *P. tremula* × *P. alba* et les gènes correspondants de *Populus trichocarpa*. La plupart des SNPs identifiés entre les espèces de peuplier sont neutres ou conduisent à des substitutions d'acides aminés conservées. Toutefois, un certain nombre de SNPs provoquent des substitutions d'acides aminés induisant des modifications biochimiques ou structurales majeures dans la protéine. En effet, l'absence ou la présence de sites putatifs de phosphorylation a peut-être des conséquences sur la régulation fonctionnelle des protéines de peuplier. Par exemple, la thréonine en position 43 de la protéine PtWLIM1a fait partie d'un

site putatif de phosphorylation par les protéines kinase C (PKC) très conservé dans la plupart des protéines WLIM1 de plantes (Fig. 34 et 45). Le remplacement de cette thréonine par une alanine dans la seconde version allélique du gène chez *Populus tremula* × *P. alba* pourrait avoir une influence sur son activité. Il est à noter que chez l'Homme, certaines mutations ponctuelles dans la protéine CRP3/MLP provoquent des pertes de fonction et sont liées à des cardiomyopathies (Knoll *et al.*, 2002 ; Mohapatra *et al.*, 2003). Ainsi, ces substitutions d'acides aminés sont nécessairement à prendre en compte lors des futures analyses fonctionnelles des protéines LIM chez le peuplier. Par ailleurs, les différents SNPs et microsatellites identifiés dans les gènes *LIM* pourront servir dans le cadre de la sélection assistée par marqueurs de peupliers hybrides, et être utilisés par les chercheurs de l'unité travaillant sur des programmes d'amélioration génétique du peuplier. Ceci sera surtout utile si ces marqueurs sont liés à des traits phénotypiques comme la qualité du bois. Par exemple, il a été montré que les variations dans l'angle des microfibrilles de bois entre les sous espèces d'eucalyptus sont associées à des SNPs portés par différents allèles d'un gène codant pour une Cinnamoyl-CoA-réductase (CCR), une enzyme de la voie de biosynthèse des lignines (Thumma *et al.*, 2005).

### 3.2. Les gènes *LIM* chez le peuplier apparaissent tous dupliqués

Le génome de *Populus trichocarpa* comprend douze modèles de gène *LIM*, alors que les génomes d'*Oryza sativa* et d'*Arabidopsis thaliana*, avec respectivement six modèles de gène *LIM*, en contiennent deux fois moins. La forte similarité de séquence observée entre les paires de gènes *LIM* de peuplier, suggère fortement une duplication récente de gènes. Ce phénomène a déjà été observé dans une étude récente concernant la famille de gène *CesA* codant pour des cellulose synthases (Djerbi *et al.*, 2005). Ces observations sont en accord avec l'hypothèse que le peuplier est un ancien polyploïde, et qu'un événement de duplication à grande échelle s'est produit chez l'ancêtre du peuplier ou au moins très tôt dans l'évolution du genre *Populus* (Sterck *et al.*, 2005 ; Tuskan *et al.*, 2006). Les analyses phylogénétiques suggèrent que ces paires de gènes *LIM* dupliqués sont des in-paralogues, c'est à dire des gènes paralogues qui ont été dupliqués après la spéciation de l'espèce et qui ont peu divergé (Sonnhammer & Koonin, 2002). En effet, aucun autre gène *LIM* provenant d'une autre espèce n'est ancré entre deux paires de gènes *LIM* de peuplier (Fig. 40). Or, il a été suggéré par des études de synténie que les deux genres *Populus* et *Salix*, de la famille des *Salicaceae*, portent

la marque de cette récente duplication à grande échelle (Tuskan *et al.*, 2006). Les gènes *LIM* dupliqués de peuplier seraient alors plutôt des out-paralogues, c'est à dire des gènes paralogues qui ont été dupliqués avant la spéciation de l'espèce, et ils pourraient aussi exister dans le génome du genre *Salix*. Toutefois, si ces gènes sont effectivement des out-paralogues, la forte identité de séquence entre ces gènes *PtLIM* dupliqués ne peut alors être expliquée que par un taux manifestement réduit dans l'évolution des séquences chez le peuplier. En effet, il a été suggéré que l'horloge moléculaire chez le peuplier est ralentie d'environ six fois par rapport à *Arabidopsis* (Tuskan *et al.*, 2006). Certains de ces gènes *LIM* dupliqués codent peut-être pour des protéines ayant gardées une fonction similaire et redondante. Or, comme nous le verrons plus tard, la distribution des ESTs ainsi que le profil d'expression entre certains gènes dupliqués diffèrent nettement, suggérant d'autres différences au sein de leurs régions régulatrices. De plus, certains de ces gènes *LIM* dupliqués ont pu acquérir de nouvelles spécificités de fonction.

### 3.3. Les gènes *LIM* de plantes ont une structure intron/exon très conservée

Les positions très conservées des quatre introns indiquent que les gènes *LIM* chez les plantes supérieures ont évolué à partir d'une séquence ancestrale commune ayant la même structure génomique (Fig. 36). De plus, le fort degré de conservation dans la longueur des régions codantes CR1, CR2, et CR4 localisées au niveau des domaines LIM, indique une stricte conservation de la longueur des domaines LIM au cours de l'évolution. Les longueurs des régions codantes reflètent les relations phylogénétiques entre les sous-groupes et groupes LIM. En effet, les protéines des groupes  $\alpha$ LIM1 et  $\beta$ LIM1 ont une deuxième région codante (CR2) de 97 paires de bases alors que les protéines des groupes  $\gamma$ LIM2 et  $\delta$ LIM2 ont une CR2 de 100 paires de bases. Enfin, au sein du groupe  $\alpha$ LIM1, la différence observée dans la longueur de la quatrième région codante des gènes *PtGLIM1a* et *b* et *OsWLIM1* illustre bien les différences phylogénétiques entre les protéines des deux sous-groupes FLIM1 et WLIM1 des monocotylédones et les protéines du sous-groupe WLIM1 des eudicotylédones.

Il a été suggéré que le gène *AtPLIM2a* ait perdu ces introns deux (I2) et quatre (I4) après la divergence entre le tournesol et *Arabidopsis* (Eliasson *et al.*, 2000). La découverte de la structure des nouveaux gènes *AtPLIM2b* et *c*, avec quatre et trois introns respectivement, nous donne plus de précisions sur la manière dont le gène *AtPLIM2a* pourrait avoir perdu ces deux introns. En effet, la longueur du second exon chez *AtPLIM2a* (129 pb) est la somme de

## Partie 1. Analyse phylogénétique

second et troisième exon de *AtPLIM2c*. Tout ceci suggère que la perte de chaque intron s'est produite séquentiellement au cours de deux événements de duplication de gènes. Premièrement, le gène ancestral *AtPLIM2b*, a pu engendrer le gène *AtPLIM2c* après la perte de l'intron quatre, et ensuite *AtPLIM2c* a pu donner *AtPLIM2a* avec la perte du deuxième intron. Enfin, la longueur des cinquièmes régions codantes, ainsi que les analyses phylogénétiques révèlent que le gène *AtPLIM2b* est plus proche des gènes *PtPLIM2a* et *b* que du gène *HaPLIM2*. De plus, l'analyse phylogénétique, nous montre que le gène *AtPLIM2b* est l'orthologue du gène *BnPLIM2a* de colza (*Brassica napus*) avec une forte similarité de séquence (89 %), et est assez éloigné des gènes *AtPLIM2a* et *c*. Enfin, il est important de signaler que le colza possède également deux gènes orthologues aux gènes *AtPLIM2a* et *c*, mais ces gènes ne sont pas représentés dans les analyses phylogénétiques, car les ESTs qui les représentent sont tronquées. De façon intéressante, ces pertes d'introns ne se sont pas produites chez les gènes *PLIM2* de peuplier et de riz, et les gènes  $\delta$ *LIM2* de peuplier plus éloignés phylogénétiquement. Tous ces éléments suggèrent donc que ces pertes d'intron semblent s'être produites spécifiquement chez les *Brassicaceae*, avant la spéciation des espèces *Arabidopsis thaliana* et *Brassica napus*.

### 3.4. L'expression des gènes *WLIM2* peut être régulée par un mécanisme d'épissage alternatif

Les gènes *WLIM2* d'eudicotylédones peuvent être régulés par des mécanismes d'épissage alternatif du premier intron, ou du deuxième intron dans le cas du gène *AtWLIM2b* ; ces deux introns étant situés dans la région 5' non codante. En effet, des variants transcriptionnels au niveau de ces introns existent pour le gène *AtWLIM2b*. De manière intéressante, ces observations sont confortées par des expériences de Northern-blot révélant dans les tiges et pas dans les racines, l'hybridation de deux bandes par une sonde *AtL2* correspondant au gène *AtWLIM2b* (Tranbarger *et al.*, 2003). Cette double hybridation ne peut correspondre à une hybridation croisée avec le gène dupliqué *AtWLIM2a*, puisque les transcrits des deux gènes *AtWLIM2a* et *b* identifiés ont approximativement la même taille (948 et 1010 pb respectivement). Enfin, trois délétions de six à huit nucléotides dans la région 5' non traduite du transcrit *PtaWLIM2a* ont été identifiées par comparaison avec la séquence génomique *PtWLIM2a*. Bien que ces deux séquences ne proviennent pas du même génotype, ces délétions pourraient correspondre à des variations d'épissage de cet intron situé dans la région 5' non codante. Ces mécanismes d'épissages alternatifs chez le gène *AtWLIM2b* et



probablement le gène *PtWLIM2a*, sont peut-être la conséquence de régulation post-transcriptionnelle de ces gènes.

### 3.5. Les résultats d'analyses phylogénétiques sont corrélés avec les données d'expression

La famille de protéine à domaine LIM avait été structurée en quatre groupes phylogénétiques : PLIM1 et PLIM2 spécifiques du pollen, et WLIM1 et WLIM2 largement exprimés dans la plante (Eliasson *et al.*, 2000). Avec les nouvelles séquences collectées chez une grande variété de plantes, la phylogénie de cette famille élargie est devenue plus complexe avec l'apparition de nouveaux groupes et sous-groupes. Toute la difficulté a été de choisir une annotation de protéine respectant à la fois la nomenclature précédemment établie, les spécificités d'expression, et les résultats obtenus par cette analyse montrant de nouvelles relations phylogénétiques entre les protéines LIM. J'ai donc défini des groupes plus larges ( $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2 et  $\delta$ LIM2) ne reflétant pas les spécificités d'expression tissulaire de ces gènes, et permettant de désigner et de classer le nouveau groupe  $\beta$ LIM1 dont les protéines sont phylogénétiquement différentes de celles des autres groupes. Toutefois, ce groupe comprend peu de protéines et ne contient apparemment pas de sous-groupes ayant des spécificités d'expression tissulaire. Les groupes PLIM1, PLIM2, WLIM1 et WLIM2 précédemment définis sont donc devenus des sous-groupes positionnés au sein des trois groupes  $\alpha$ LIM1,  $\gamma$ LIM2 et  $\delta$ LIM2. De plus, de nouveaux sous-groupes comme FLIM1 ont pu ainsi être défini. Enfin, cette classification permet à certaines protéines difficiles à positionner dans un sous-groupe particulier d'être nommées avec un préfixe d'alphabet grec, quitte à être renommée plus tard quand leurs profils d'expression seront connus. De plus, la classification phylogénétique de cette famille sera, bien sûr, améliorée lorsque plus de séquences codant pour des protéines à domaine LIM auront été identifiées. Enfin, il serait intéressant pour les 51 autres unigènes identifiés, mais codant pour des protéines LIM partielles, d'obtenir des ADNc pleine longueur afin de compléter ce répertoire de gène *LIM*.

#### 3.5.1. Le groupe $\alpha$ LIM1 hétérogène, comporte des sous-groupes avec des spécificités d'expression.

Les résultats d'analyse phylogénétique montrent que le groupe  $\alpha$ LIM1 contient beaucoup de protéines et est un groupe hétérogène comprenant plusieurs sous-groupes bien

## Partie 1. Analyse phylogénétique

séparés (PLIM1, FLIM1, WLIM1 des dicotylédones et WLIM1 des monocotylédones). Par contre, au sein de ce groupe  $\alpha$ LIM1, les origines de cette diversification structurale et de ces spéciations sont incertaines (Fig. 40 et 41). Ainsi, on ne peut pas déterminer quel est le sous-groupe apparu en premier, et il semblerait que ces quatre sous-groupes ont pu apparaître indépendamment au cours de l'évolution. Afin d'affiner l'organisation phylogénétique de ces différents sous-groupes au sein du groupe  $\alpha$ LIM1, plus de taxons sont nécessaires, notamment des séquences protéiques LIM provenant d'espèces ancestrales de plante comme les fougères ou d'autres espèces de la classe des magnolidées.

Le sous-groupe FLIM1 comprend les protéines PtGLIM1a et b de peuplier et les protéines  $\alpha$ LIM1 d'euphorbe, de ricin et de cotonnier. Bien que des séquences additionnelles soient nécessaires pour consolider cette classification, il y a de forts arguments pour la création d'un sous-groupe FLIM1 : i) Les protéines du sous-groupe FLIM1 présentent une forte identité de séquence entre elles avoisinant 80 % et allant jusqu'à 87 % entre les protéines de peuplier et celle de ricin. ii) Comme nous allons le voir plus tard, les analyses RT-PCR indiquent que les gènes *PtGLIM1a* et *b* ne sont pas exprimés dans le pollen mais plutôt dans les tissus vasculaires. Ainsi ces deux gènes *PtGLIM1a* et *b* peuvent difficilement être groupés au sein du sous-groupe PLIM1 spécifique du pollen. De la même manière, les 31 ESTs *GhaLIM1* ont toutes été trouvées dans les fibres de coton matures, et une majorité (28 ESTs) proviennent de fibres récoltées 20 jours après l'anthèse, à un stade de développement correspondant à la formation de la paroi secondaire (Annexe 2). J'ai donc nommé ce sous-groupe FLIM1 avec le préfixe F pour la forte expression dans les fibres de coton ou les fibres de bois de tension chez le peuplier. Nous pouvons spéculer que durant l'évolution des espèces de peupliers et de cotonniers, ces trois protéines ont pu gagner une nouvelle fonction importante pour la différenciation ou la maturation des fibres et plus particulièrement des fibres contenant une paroi cellulaire secondaire enrichie en cellulose. Par contre, pour les deux autres gènes *EeaLIM1* et *RcaLIM1*, nous ne pouvons pas spéculer sur leurs profils d'expression étant donné qu'une seule EST n'a été isolée pour chacun de ces gènes. Enfin, la protéine Cs $\alpha$ LIM1 de l'oranger est difficile à positionner au sein des sous-groupes FLIM1 ou WLIM1 des eudicotylédones. En effet, il semblerait que celle-ci soit phylogénétiquement proche du sous-groupe FLIM1, mais la répartition des 28 ESTs identifiées suggère que ce gène est largement exprimé dans tous les tissus de la plante.

Beaucoup d'études d'expression ont démontré l'expression spécifique dans le pollen des gènes *PLIM1*. Par exemple, les transcrits *HaPLIM1a* (*SF3*) représentent environ 3 % des ARNm totaux de pollen (Baltz *et al.*, 1992a), et la protéine HaPLIM1a est exclusivement et

fortement exprimée dans le pollen mature (Baltz *et al.*, 1999). De plus, sa co-localisation avec l'actine dans les cônes germinatifs de pollen suggère que la protéine HaPLIM1a participe dans la germination du pollen ou l'élongation du tube pollinique. Toutefois, il est étonnant de n'avoir collecté aucune EST correspondant au gène *HaPLIM1a* (Annexe 2). Par contre, 23 ESTs homologues au gène *HaPLIM1b* ont été trouvées. Le gène *HaPLIM1a* est peut-être un allèle du gène *HaPLIM1b*, et ne résulte probablement pas d'une duplication de gène. De plus, aucun gène dupliqué *PLIM1* n'a été trouvé pour les autres espèces de la famille des astéracées. Les gènes *NtPLIMa* et *b* de tabac sont également fortement exprimés dans le pollen mature (Sweetman *et al.*, 2000), et chez *Gerbera hybrida*, des études microarray montrent que le gène *GehPLIM1a* est plus de huit fois surexprimé dans les étamines par rapport aux autres organes (Laitinen *et al.*, 2005). De la même manière, une autre étude indique que le gène *PGPS/D1* de pétunia codant une protéine PLIM1 est fortement exprimé dans les pollens en germination (Guyon *et al.*, 2000). L'accumulation du transcrit *PhPLIM1* augmente avec la maturation des anthères et continue de s'accumuler deux heures après la germination suggérant encore une implication des protéines PLIM1 dans l'élongation du tube pollinique ou dans les interactions avec le pistil. La particularité du sous-groupe PLIM1 spécifique du pollen est de contenir uniquement des protéines issues d'espèces des familles de *Solanaceae* et *Asteraceae* de la sous-classe des astéridées (Fig. 40). Cette particularité montre que les protéines PLIM1 sont sûrement apparues durant la spéciation de ces deux familles de plantes au cours de l'évolution. Ces protéines PLIM1 ont divergées et se sont spécialisées dans des fonctions importantes pour le développement du pollen comme la germination ou l'élongation du tube pollinique. Ces fonctions polliniques particulières, non acquises chez les espèces de rosidées et monocotylédones, et spécifiques des espèces de la sous-classe des astéridées, ont été conservées chez ces espèces car elles leurs confèrent probablement un avantage sélectif lors de la pollinisation ou de la fécondation. De même, on constate que les protéines PLIM1 des espèces de la famille des *Solanaceae* sont nettement différentes et phylogénétiquement séparées de celles des espèces de la famille des *Asteraceae*, suggérant que la fonction des protéines PLIM1 n'est pas la même entre les familles de *Solanaceae* et d'*Asteraceae*. Ces deux types de protéines PLIM1 ont divergé durant la spéciation de ces deux familles et peuvent ainsi participer à des fonctions spécifiques impliquées dans des processus reproductifs distinguant les espèces de la famille des *Solanaceae* de celles de la famille des *Asteraceae*.

Le sous-groupe WLIM1 des monocotylédones a également été appelé WLIM1, car les EST correspondantes proviennent d'une grande variété de tissus pour la plupart végétatifs,

mais aussi reproducteurs (Annexe 2). De façon surprenante, un grand nombre d'ESTs WLIM1 ont été séquencées à partir de banques d'ADNc de maïs (200 ESTs *ZmWLIM1*) et de canne à sucre (128 ESTs *SoWLIM1a* et *b*). En particulier, le profil d'expression des gènes *SoWLIM1a* et *b* est intéressant : Le gène *SoWLIM1a*, avec 54 ESTs, est exprimé dans tous les tissus, surtout la zone de transition tige/racine, les inflorescences à tous les stades de développement, et les méristèmes apicaux. Au contraire, le gène *SoWLIM1b*, avec 74 ESTs (sur 75 identifiées), est le plus gros cluster de la banque d'ADNc de « la zone de transition racine/tige de plante adulte », suggérant que ce gène est fonctionnellement important dans cette zone particulière chez les monocotylédones. Jusqu'à présent, peu d'études font référence à cette zone de transition entre les tissus vasculaires de la tige et ceux de la racine. Cette zone de transition correspond à une zone de changement dans l'organisation des tissus vasculaires chez les monocotylédones (Ye, 2002) : Dans la racine, les paquets vasculaires sont organisés en cercle (eustèle) alors que dans la tige, ils sont éparpillés d'un bout à l'autre des tissus (atactostèle). Ainsi, les protéines WLIM1 interviennent peut-être dans l'organisation et/ou la connexion de ces deux types d'éléments vasculaires. Chez le riz, cette zone de transition racine/tige dépourvue d'aérenchyme est la principale zone de résistance au transport de l'oxygène des tiges vers les racines (Groot *et al.*, 2005). De façon plus intéressante, il a été montré que l'acide 1-N-Naphthylphthalamique (NPA) utilisé comme herbicide, et les flavonols tels que le kaempférol et la quercétine s'accumulent dans la zone de transition racine/hypocotyle de jeunes plantules d'*Arabidopsis* (Murphy & Taiz, 1999 ; Peer *et al.*, 2001). Sous l'action du NPA, un inhibiteur potentiel du transport polaire de l'auxine, cette hormone s'accumule fortement dans cette zone (Sabatini *et al.*, 1999), et les flavonols jouent également un rôle dans l'inhibition du transport de l'auxine (Brown *et al.*, 2001 ; Besseau *et al.*, 2007). Cette zone particulière de transition entre les racines et les tiges pourrait donc agir comme un point de contrôle du transport de l'auxine. Enfin, la forte conservation entre les séquences des protéines WLIM1 de ces différentes espèces de monocotylédones est un indicateur de leur probable implication dans des fonctions essentielles pour la plante.

En général, chaque gène *WLIM1* d'eudicotylédones est représenté par un grand nombre d'ESTs originaires de tissus très variés, incluant parfois les inflorescences. Les gènes *WLIM1* sont les gènes *LIM* les plus étudiés chez les plantes, mais leurs fonctions précises restent indéfinies. Chez *Arabidopsis*, le gène *AtWLIM1* est légèrement plus exprimé dans les tiges que dans les feuilles et les fleurs. Son orthologue, le gène *BnWLIM1* (ou *BnLIM1*) chez le colza (*Brassica napus*), est aussi exprimé dans les organes végétatifs et reproducteurs (Huang *et al.*, 2007). De plus, son transcrit s'accumule durant le remplissage des graines et

## Partie 1. Analyse phylogénétique

l'expression du gène dans les plantules est induite par l'ajout de saccharose dans le milieu. Comme le profil d'expression du gène *BnWLIM1* est similaire à celui du gène *CCR*, les auteurs suggèrent que *BnWLIM1* joue un rôle dans la régulation de la biosynthèse des lignines. Enfin, la fusion de la protéine *BnWLIM1* avec la GFP montre une localisation de la protéine dans le noyau et le cytoplasme de cellules épidermiques d'oignon (Huang *et al.*, 2007). Chez le tournesol, la protéine *HaWLIM1* est aussi localisée à la fois dans le noyau et le cytoplasme pour différents types cellulaires (Mundel *et al.*, 2000), et pourrait être associée dans les protoplastes avec le cytosquelette de microtubules (Brière *et al.*, 2003). Le gène *WLIM1* de tabac, *NtLIM1*, plus exprimé dans les tiges que dans les feuilles et les racines, agit comme un facteur de transcription. En effet, la fixation de la protéine *NtLIM1* au motif de la boîte PAL (PAL-box), présent au niveau des promoteurs de plusieurs gènes impliqués dans la voie de biosynthèse des lignines, conduit à l'activation de la transcription de ces gènes, en l'occurrence les gènes *PAL*, *4CL* et *CAD*. De plus, une réduction importante du contenu en lignine a été observée dans des tabacs transgéniques antisens *Ntlim1* (Kawaoka *et al.*, 2000). Chez l'eucalyptus (*Eucalyptus globulus* et *E. camaldulensis*) les gènes et ADNc *EgLIM1* et *EcLIM1* ont été isolés. La réduction de l'expression du gène *EcLIM1* semble aussi diminuer l'expression des gènes *PAL*, *4CL* et *CAD* et la teneur en lignine dans les plantes transgéniques portant une construction antisens *Ntlim1* (Kawaoka *et al.*, 2006). Comme chez le tournesol, la protéine *NtWLIM1* de tabac est à la fois localisée dans le cytoplasme et le noyau (Kawaoka & Ebinuma, 2001 ; Thomas *et al.*, 2006). Dans le cytoplasme, la protéine *NtWLIM1* lie directement l'actine avec une forte affinité, augmente la stabilité du cytosquelette d'actine et favorise l'empaquetage des filaments d'actine (Thomas *et al.*, 2006). Cette probable double fonction cellulaire des protéines *WLIM1* comme facteur de transcription et comme protéine se liant à l'actine est intrigante quand au type de processus physiologique pour lequel ces protéines sont impliquées dans la plante. Néanmoins, la forte conservation de séquence entre les protéines *WLIM1* d'eudicotylédones indique sûrement un rôle dans des fonctions cellulaires indispensables à la plante.

### 3.5.2. Les protéines LIM1 de conifères et de fougères sont ancestrales aux groupes $\alpha$ LIM1 et $\beta$ LIM1

Les analyses phylogénétiques placent les protéines LIM1 de conifères et de fougères à l'intérieur ou à l'extérieur du groupe  $\beta$ LIM1 selon la méthode utilisée (Fig. 39, Annexes 4 et 5). Il paraît toutefois certain, de part les fortes valeurs de bootstrap, que ces protéines sont

## Partie 1. Analyse phylogénétique

phylogénétiquement séparées des protéines des groupes  $\alpha$ LIM1  $\gamma$ LIM2 et  $\delta$ LIM2. Ces protéines LIM1 de conifères et de fougères peuvent être considérées comme les ancêtres communs aux protéines du groupe LIM1 de plante, et probablement aux protéines  $\beta$ LIM1. Ce groupe  $\beta$ LIM1 serait alors apparu en premier au cours de l'évolution des protéines LIM1 des angiospermes. Cependant, cette hypothèse reste à être vérifiée par l'ajout dans cette analyse de séquences  $\beta$ LIM1 supplémentaires, provenant d'espèces ancestrales aux angiospermes comme les fougères, les algues ou les lichens. Un aspect intéressant est la forte conservation de séquence entre les protéines LIM1 de conifères, surtout avec 98 % d'identité entre les protéines LIM1b de *Picea glauca* et *Pinus taeda*, deux espèces de genre différent. Cette forte conservation de séquence contraste avec le polymorphisme de substitution en acides aminés observé dans le genre *Populus*. Il a en effet été observé que chez *Pinus sylvestris* et *P. taeda*, le niveau de polymorphisme silencieux est faible comparé aux autres espèces de plante (Dvornyk *et al.*, 2002 ; Neale & Savolainen, 2004). Chez l'épicéa (*Picea glauca*), les ESTs *PgLIM1a* et *b* sont retrouvées dans tous les tissus incluant les aiguilles et les cônes, mais c'est dans les racines et les tissus vasculaires du tronc (phloème, cambium et xylème) que l'on trouve le plus d'ESTs (Annexe 2). Chez le pin à encens (*Pinus taeda*), le gène *PitLIM1a* contient deux tiers de ces ESTs provenant de racines, et le reste de ces ESTs proviennent majoritairement de xylème de tige. D'autre part, des expériences de Northern-blot montrent que le gène *PtLIM1b* (clone 6C12H) est plus fortement exprimé dans le xylème immature à paroi primaire que dans d'autres parties de l'arbre telles que les aiguilles (Allona *et al.*, 1998). De plus, l'expression du gène est moins forte dans le xylème mature ayant formé de la paroi secondaire, ainsi que dans le phloème. Au niveau d'un pin incliné, le gène est plus faiblement exprimé dans le xylème immature de bois de compression que dans celui de bois « normal » prélevé sur les cotés du tronc incliné. A l'inverse du bois de tension chez les angiospermes, le bois de compression des gymnospermes contient plus de lignine et moins de cellulose que le bois normal. Ainsi, ce résultat d'expression n'est pas en accord avec l'hypothèse selon laquelle les protéines LIM interviendraient dans l'activation de la biosynthèse des lignines, comme il a été suggéré pour le facteur de transcription NtLIM1 de tabac (Kawaoka *et al.*, 2000). Toutefois, au moins chez les conifères, les gènes LIM1 seraient peut-être préférentiellement exprimés durant le développement vasculaire.

3.5.3. Le nouveau groupe  $\beta$ LIM1 est encore à explorer

De façon intéressante, au sein du groupe  $\beta$ LIM1, aucune protéine d'*Arabidopsis* ni d'espèces de la famille des *Brassicaceae* n'a été identifiée (Fig. 42). De la même manière, parmi les monocotylédones, aucune protéine d'espèces de la famille des *Poaceae* n'est présente dans ce groupe. Par contre, nous trouvons dans ce groupe  $\beta$ LIM1 la protéine Zo $\beta$ LIM1 appartenant à la famille des *Zingiberaceae*. Il est important de signaler que le gène *OsLIM* codant pour une protéine à deux domaines LIM inhabituelle est plutôt proche des protéines  $\beta$ LIM1, si l'on ne considère pas son long domaine C-terminal. La protéine la plus proche de *OsLIM* est Zo $\beta$ LIM1, mais le pourcentage d'identité en acides aminés n'est pas très élevé (60 %). Par ailleurs, j'ai identifié une autre protéine appartenant à l'espèce *Festuca arundinacea* de la famille des *Poaceae* (non représenté dans l'analyse phylogénétique car tronquée), très proche de *OsLIM* avec 87 % d'identité au niveau des domaines LIM. Il serait intéressant de vérifier si cet ADNc *OsLIM* correspond bien à une réalité biologique et n'est pas la conséquence d'une erreur de clonage. Par ailleurs, l'absence de protéines d'*Arabidopsis* dans ce groupe  $\beta$ LIM1 est difficile à expliquer puisque l'expression ou la fonction des gènes  $\beta$ LIM1 de plante est encore inconnue. Néanmoins, le nouveau groupe  $\beta$ LIM1 n'est vraisemblablement pas spécifique d'une classe ou sous-classe de plante supérieure. Comme les gènes des sous-groupes WLIM1 et WLIM2, les ESTs  $\beta$ LIM1 sont originaires d'une grande variété de tissus incluant les tiges, les racines, les feuilles, les fleurs et les fruits (Annexe 2). Beaucoup d'ESTs ont été préférentiellement trouvées dans les racines et les cellules en cours de différenciation comme les fibres de coton immatures et des cellules indifférenciées tels que les cals cellulaires. Comme nous le verrons plus tard, les ESTs *Pt $\beta$ LIM1a* proviennent en majorité de cellules en culture, alors que les ESTs *Pt $\beta$ LIM1b* proviennent de tissus différents. Contrairement aux gènes *PtGLIM1a* et *b* et *PtWLIM1a* et *b*, ces gènes ne semblent pas être préférentiellement exprimés dans les tissus vasculaires du tronc. Bien que les ESTs  $\beta$ LIM1 semblent provenir de toutes les parties de la plante, un nombre significatif d'ESTs a été séquencé à partir d'échantillons de cellules indifférenciées. Il semblerait donc que ces protéines  $\beta$ LIM1 soient plutôt impliquées dans la différenciation cellulaire ou le maintien de l'état indifférencié des cellules, mais des études d'expression plus précises sont nécessaires pour affiner l'étude fonctionnelle des protéines  $\beta$ LIM1.

### 3.5.4. Les gènes du groupe $\gamma$ LIM2 important pour la plante seraient probablement régulés par des mécanismes ressemblant à de l'épigénétique

Pour les deux sous-groupes WLIM2 de eudicotylédones et monocotylédones, un grand nombre d'ESTs ont été séquencées à partir de tissus variés, incluant les inflorescences (Annexe 2). Notamment, les gènes *AtWLIM2b* d'*Arabidopsis*, *TaWLIM2* de blé, *ZmWLIM2* de maïs, et *PgWLIM2* d'épicéa avec respectivement 159, 134, 86 et 47 ESTs identifiées semblent être largement exprimés dans la plante. Les gènes *NtWLIM2* de tabac et *AtWLIM2a* d'*Arabidopsis* sont tous deux exprimés dans les feuilles, tiges et fleurs ; et bien que *NtWLIM2* soit exprimé dans les anthères, son expression n'est pas détectée dans le pollen (Eliasson *et al.*, 2000). Chez le coton, des analyses d'expression montrent que le gène *GhWLIM2a* (ou *GhLIMI*) est exprimé dans une grande variété de tissus, avec une expression préférentielle dans les bourgeons apicaux, les fibres de coton et les ovules. Il a été proposé que l'expression du gène *GhWLIM2a* soit liée au développement de la fibre de coton (Luo *et al.*, 2003). Le vaste éventail de tissus dont sont originaires ces ESTs, ainsi que le grand nombre d'EST identifiées et la forte conservation de séquence entre les protéines WLIM2, suggèrent que ces gènes *WLIM2* sont importants dans des processus cellulaires de base. Ceci expliquerait peut-être la difficulté d'obtenir des mutants homozygotes pour le gène *AtWLIM2a* (Dean Rider *et al.*, 2003). En effet, lors de la ségrégation de plantes hétérozygotes pour un allèle d'insertion ADN-T du gène *AtWLIM2a* certaines présentent un phénotype létal au stade embryonnaire. Bien qu'il n'ait pas été déterminé dans cette étude si ce phénotype est lié à l'allèle mutant *Atwlim2a*, ce résultat préliminaire suggère que le gène *AtWLIM2b* ne compenserait pas la perte de fonction du gène *AtWLIM2a*. Dans une étude de microarray et RT-PCR quantitative, ce gène *AtWLIM2a* apparaît fortement exprimé dans les siliques d'*Arabidopsis*, et son expression est induite dans les graines en germination du mutant *pkl* d'*Arabidopsis* (Dean Rider *et al.*, 2003). La protéine PICKLE (PKL), code pour un facteur putatif CHD3 de remodelage de la chromatine modifiant le degré d'acétylation des histones. PKL est un régulateur négatif de la transcription qui agit comme un répresseur de l'identité embryonnaire durant la germination des graines. De manière analogue au gène *LEAFY COTYLEDON1* (*LEC1*), un régulateur central de l'embryogenèse, le gène *AtWLIM2a* pourrait ainsi être impliqué dans la régulation du développement de l'embryon. Dans une autre étude d'expression chez *Arabidopsis*, l'expression du gène *AtWLIM2a* est induite dans les feuilles en sénescence (Gepstein *et al.*, 2003). Plus précisément, cette induction est bi phasique dans le temps, suggérant une fonction à la fois durant la maturation de la feuille et plus tard



seulement au stade terminal du programme de sénescence de la feuille (Gepstein *et al.*, 2003). Ces deux résultats pris ensemble suggèrent que le gène *AtWLIM2a* pourrait être plus précisément impliqué dans la maturation des graines ou dans la réponse à l'éthylène. Le gène dupliqué *AtWLIM2b* (*AtL2*), est plus fortement exprimé dans les racines que dans les tiges d'*Arabidopsis*. De plus, son expression semble être affectée par la disponibilité en azote (Tranbarger *et al.*, 2003). De manière intéressante, chez le palmier à huile (*Elaeis guineensis*), l'ADNc *EgWLIM2* a été utilisé comme marqueur RFLP (sonde CPHO63) sensible à la méthylation de l'ADN, pour identifier un polymorphisme d'origine épigénétique associé à des variations somaclonales (Jaligot *et al.*, 2002). Ce marqueur est capable de différencier des cals à croissance normale de cals à croissance rapide, générant respectivement 5 % et 100 % de plantules « *mantled* ». L'anomalie « *mantled* » est caractérisée chez le palmier adulte par une structure florale aberrante avec les parties mâles transformées en pseudo carpelles (Rival *et al.*, 1997). Bien que les changements observés dans la méthylation de l'ADN ne reflètent pas forcément des modifications spécifiques de l'activité du gène *EgWLIM2*, il est intéressant de constater que l'expression des deux gènes *EgWLIM2* et *AtWLIM2a* pourraient être régulée par des modifications épigénétiques résultant respectivement de la méthylation de l'ADN et du remodelage de la chromatine par la protéine PKL. De plus, ces deux gènes *EgWLIM2* et *AtWLIM2a* interviendraient dans la différenciation de cellules d'un état embryonnaire ou indifférencié (cals) à un état de plantule.

### 3.5.5. Les gènes du groupe $\delta$ LIM2 sont préférentiellement exprimés dans le pollen

Le groupe  $\delta$ LIM2 comprend des protéines très divergentes entre elles du fait de leurs domaines C-terminal atypiques. Elles ont été classées en trois sous-groupes : les sous-groupes PLIM2 des monocotylédones et eudicotylédones et  $\delta$ LIM2 des astéridées (Fig. 44). La plupart des gènes *PLIM2* des eudicotylédones apparaissent préférentiellement exprimés pendant le développement du pollen. Chez *Arabidopsis*, le gène *AtPLIM2a* est plus fortement exprimé dans les grains de pollen hydratés que dans les organes végétatifs (Eliasson *et al.*, 2000 ; Becker *et al.*, 2003). De plus, les résultats d'expression fournis par la base de donnée d'analyses microarray « NASCArrays » (Craigon *et al.*, 2004) hébergée sur le site (<http://affymetrix.arabidopsis.info/>), nous indiquent que les trois gènes *AtPLIM2a*, *b* et *c* sont dix fois plus exprimés dans les anthères matures et le pollen que dans les anthères à un stade précoce de développement. C'est aussi le cas du gène *NtPLIM2* de tabac uniquement exprimé

## Partie 1. Analyse phylogénétique

dans les anthères matures et les grains de pollen, et du gène *HaPLIM2* de tournesol également exprimé dans les grains de pollen (Eliasson *et al.*, 2000). De la même manière, chez *Gerberaea hybrida*, le gène *GhPLIM2* orthologue au gène *HaPLIM2* apparaît exclusivement exprimé dans les étamines (Laitinen *et al.*, 2005), et chez la tomate (*Lycopersicon pennellii*), toutes les ESTs *LpPLIM2* orthologues au gènes *NtPLIM2* sont originaires d'une banque d'ADNc de pollen (*GhPLIM2* et *LpPLIM2* sont deux séquences incomplètes non montrées dans l'analyse phylogénétique). Enfin, comme nous le verrons plus tard, les gènes *PtPLIM2a* et *b* sont tous deux fortement exprimés dans les étamines et le pollen de peuplier, et leurs transcrits sont aussi détectés dans d'autres tissus tel que le xylème.

Comme pour les eudicotylédones, les gènes *PLIM2* des monocotylédones sont la plupart du temps fortement exprimés dans le pollen, mais pas exclusivement. 156 ESTs parmi les 200 ESTs *ZmPLIM2a* trouvées dans les bases de données publiques proviennent de banques d'ADNc de pollen, et le reste de ces ESTs sont originaires de mélanges de tissus non caractérisés (Annexe 2). De même, chez le blé, la majorité des ESTs *TaPLIM2b*, et toutes les ESTs *TaPLIM2a* et *c* proviennent de banques d'ADNc construites à partir d'épis et d'inflorescences. Pour le riz, la majorité des ESTs *OsPLIM2c* et particulièrement les ESTs *OsPLIM2b* viennent aussi de banques d'ADNc de fleurs. L'expression du gène *OsPLIM2c* (AK072520) de riz a été étudiée par deux analyses microarray différentes. Au cours du développement des anthères, le transcrit *OsPLIM2c* s'accumule fortement dans les anthères au stade mature (Endo *et al.*, 2004). Durant la pollinisation du pistil jusqu'à l'embryogenèse précoce, son accumulation maximale dans le pistil apparaît à l'anthèse et l'expression du gène décroît graduellement jusqu'à 24 heures après l'anthèse quand le tube pollinique atteint le micropyle (Yoshida *et al.*, 2005). Toutefois, la même étude révèle une expression plus large du gène *OsPLIM2c* dans différents tissus suggérant une implication dans d'autres processus que l'élongation du tube pollinique. Ainsi, les gènes *PLIM2* de monocotylédones et d'eudicotylédones semblent tous très fortement exprimés dans le pollen, mais de façon préférentielle. Nous avons vu qu'une spécialisation de gènes appartenant au groupe  $\alpha$ LIM1 vers une expression spécifique du pollen à travers le sous-groupe PLIM1 ne s'est pas produite chez les espèces de la classe des liliopsidées et de la sous-classe des rosidées. Nous pouvons spéculer qu'en contrepartie, les gènes *PLIM2* de ces espèces se sont largement dupliqués au sein du groupe  $\delta$ LIM2 augmentant le nombre de gènes impliqués dans des fonctions polliniques. Ceux-ci pourraient ainsi compenser le manque de spécialisation des gènes du groupe  $\alpha$ LIM1 vers une expression spécifique du pollen. Cette large duplication de gènes a pu

également permettre à certains gènes *PLIM2* dont l'expression n'est pas exclusive du pollen, d'acquies d'autres fonctions non liées à la croissance du tube pollinique.

Le sous-groupe  $\delta$ LIM2 des astéridées est très proche des sous-groupes PLIM2 des eudicotylédones et monocotylédones et contient des protéines très divergentes entre elles possédant, comme les protéines PLIM2, un long domaine C-terminal à caractère acide (Fig. 45 et Annexe 6). Nous n'avons malheureusement pas à notre disposition de données d'expression concernant ces gènes. De plus, de par le faible nombre d'ESTs séquencées pour ces gènes, les données de répartition d'ESTs ne sont pas assez robustes. Toutefois, les quatre ESTs *Lp $\delta$ LIM2* de tomate identifiées proviennent toutes de banques d'ADNc de pollen. De plus, comme il apparaît évident que les gènes *PLIM2* des eudicotylédones et des monocotylédones sont tous préférentiellement exprimés dans le pollen, il est probable que ces gènes  $\delta$ LIM2 des astéridées soient aussi fortement exprimés dans le pollen. Cependant, il n'est pas certain que ces gènes *PLIM2* et  $\delta$ LIM2 aient exactement la même fonction, et cette duplication plutôt ancienne au cours de l'évolution a pu apporter une nouvelle fonction impliquée dans le développement du pollen ou dans d'autres processus. Par ailleurs, il est important de mentionner dès à présent que les deux gènes *Pt $\delta$ LIM2a* et *b*, très proches phylogénétiquement des gènes  $\delta$ LIM2 des astéridées, sont aussi fortement exprimés dans le pollen comme les gènes *PtPLIM2a* et *b*.

### 3.6. Vers une nouvelle caractérisation des séquences consensus des deux domaines LIM chez les plantes

Les protéines à domaine LIM chez les plantes possèdent un second domaine LIM particulier comparé aux protéines CRP animales. En effet, la seconde cystéine du deuxième doigt de zinc est remplacée par une glycine conservée. Il a été proposé que l'un des deux acides aminés proches de cette glycine, une histidine ou une cystéine, permette la formation de deux structures alternatives -C-X-H-X<sub>18</sub>-C-X<sub>2</sub>-H- ou -C-X<sub>4</sub>-C-X<sub>15</sub>-C-X<sub>2</sub>-H- pour la coordination de l'atome de zinc (Eliasson *et al.*, 2000 ; Mundel *et al.*, 2000). Toutefois, l'augmentation du nombre de protéines analysées révèle que le résidu histidine est quelque fois manquant dans les protéines à domaine LIM issues des groupes  $\beta$ LIM1 et  $\delta$ LIM2 (Annexe 6). Seul, le résidu cystéine est strictement conservé pour toutes les protéines à domaine LIM de plante, suggérant fortement que la structure -C-X<sub>4</sub>-C-X<sub>15</sub>-C-X<sub>2</sub>-H- est la plus probable pour la formation du second doigt de zinc. Par ailleurs, nous avons vu que la première cystéine du premier domaine LIM est manquante pour la protéine SoWLIM1b et la

seconde version allélique de *ZmWLIM1* chez la canne à sucre et le maïs, deux espèces de la même tribu des *Andropogoneae*, donc très proches phylogénétiquement. Si cette cystéine est nécessaire à la formation du doigt de zinc, la signification fonctionnelle de cette absence reste à être élucidée car une telle mutation engendrerait probablement une perte de fonction. En effet, l'arginine remplaçant la cystéine ne semble pas être un acide aminé capable de lier le zinc (Kadrmas & Beckerle, 2004), et en conséquence les structures secondaire et tertiaire du domaine LIM1 seraient affectées par une telle mutation. Dans l'hypothèse où ce type de protéine WLIM1 serait non fonctionnel, il est difficile d'expliquer sa conservation dans la plante au cours de l'évolution. Toutefois, la répartition des ESTs *SoWLIM1b* nous montre nettement que ce gène est préférentiellement exprimé dans la zone de transition tige/racine de la plante, suggérant l'acquisition d'une nouvelle fonction impliquée dans cette zone particulière chez les graminées. Nous pouvons tout de même signaler que, au moins pour la canne à sucre, le gène *SoWLIM1a* contenant cette cystéine est présent pour assurer la fonction WLIM1 dans la plante. Cette particularité des protéines WLIM1 chez la canne à sucre ouvre des perspectives de recherche attirantes, et pourra être utilisée pour mieux comprendre la structure et la fonction des domaines LIM chez les plantes. A l'intérieur du premier domaine LIM, dans la boucle du premier doigt de zinc, le motif K[T/A]VY est très conservé à la fois pour les protéines à domaine LIM de plante (Fig. 45) et les protéines CRP animales. Ce motif commun aux deux règnes animal et végétal, a sûrement une importance dans la bonne conformation des domaines LIM1 et LIM2. Toutefois ce motif est moins conservé au niveau du domaine LIM2. Enfin, ce motif participe pour les protéines du groupe  $\beta$ LIM1, du sous-groupe WLIM2 et certaines protéines  $\delta$ LIM2 à la formation d'un site putatif de phosphorylation sur la tyrosine, indiquant une probable importance de ce motif K[T/A]VY dans la régulation de l'activité de ces protéines.

Les protéines à domaine LIM sont très conservées entre elles, notamment les protéines des sous-groupes WLIM1 et WLIM2 et du groupe  $\beta$ LIM1. Cette conservation de séquence est principalement due à l'existence des deux domaines LIM très conservés d'interaction entre protéines et/ou entre protéine et ADN. De même, l'actine est une autre protéine très conservée qui évolue très peu même entre les différents règnes (Doolittle, 1992). En effet, c'est une protéine fondamentale pour l'organisation cellulaire des eucaryotes et elle a un taux de substitution en acide aminé très faible comparable à celui des histones et de la tubuline. De plus les procaryotes ne contiennent pas d'actine ni de protéines à domaine LIM (Kadrmas & Beckerle, 2004). Dans l'hypothèse où les protéines à domaine LIM interagiraient toutes avec l'actine, comme cela a été montré pour la protéine NtWLIM1 de tabac (Thomas *et al.*, 2006),

il n'est alors pas surprenant de constater une telle conservation de séquence au niveau des domaines LIM de ces protéines. Cette conservation de séquence serait le résultat d'une co-évolution entre les protéines LIM et son partenaire l'actine.

### *3.7. Les variations structurelles des protéines à domaine LIM peuvent refléter de différences fonctionnelles*

Comme chez les animaux, les protéines à domaine LIM de plante peuvent avoir une fonction dans la régulation de la différenciation ou du développement des cellules. Les analyses phylogénétiques montrent que les protéines LIM ont divergées entre elles pour former des groupes et sous-groupes LIM spécialisés dans des fonctions particulières aux plantes, comme la formation du tube pollinique, le développement des fibres de coton, la différenciation de cellules embryonnaires, le développement et la différenciation vasculaire, et la formation de bois de tension. Ainsi, des motifs d'acides aminés conservés, et spécifiques de ces groupes et sous-groupes soulignent probablement les différences fonctionnelles entre ces protéines. Avec la grande quantité de séquences récupérées, les comparaisons de séquences orthologues ont été facilitées ainsi que la découverte de motifs fortement conservés. En effet, les protéines orthologues sont en général plus conservées dans leurs séquences que les protéines paralogues et la recherche de motifs peptidiques conservés ayant une fonction potentielle dans les familles de protéines est plus puissante quand on utilise un alignement d'un ensemble de protéines orthologues plutôt qu'un ensemble de protéines paralogues (Panstruga, 2005). Or on peut noter qu'au sein des sous-groupes de protéines orthologues, on a des protéines non orthologues entre elles comme les paires de protéines LIM dupliquées de peuplier ou les protéines AtPLIM2a, b et c d'*Arabidopsis*. Ces protéines seraient plutôt considérées comme des in-paralogues (ou co-orthologues) où la duplication des gènes s'est effectuée au sein de l'espèce, ou bien des out-paralogues pour lesquelles la duplication du gène s'est produite chez un ancêtre commun (Sonnhammer & Koonin, 2002) qui serait dans le cas particulier du peuplier très proche phylogénétiquement (Tuskan *et al.*, 2006). Néanmoins, il est très vraisemblable que ces protéines in ou out-paralogues puissent avoir une fonction très proche.

La présence de sites de phosphorylation ou glycosylation très conservés entre les différentes espèces de plantes analysées, suggère une fonction clé pour ces sites. Notamment, le site de phosphorylation PKC trouvé dans le domaine C-terminal des protéines du sous-groupe WLIM2 a été conservé au cours de l'évolution pour toutes les protéines WLIM2,

## Partie 1. Analyse phylogénétique

incluant même les protéines WLIM2 de conifères. Il est probable qu'une mutation intervenant au niveau de ce site de phosphorylation conduise à une perte de fonction, et ainsi une telle mutation pourrait être létale pour la descendance. Enfin, les différences observées entre les protéines des différents sous-groupes LIM au niveau des sites putatifs de phosphorylation ou glycosylation expliquent probablement leurs différences fonctionnelles. L'absence du site de phosphorylation PKC au niveau du domaine N-terminal des protéines du sous-groupe FLIM1 est un exemple des différences observées entre ces protéines et les autres protéines des sous-groupes WLIM1, PLIM1 et du groupe  $\beta$ LIM1. La perte de ce site de phosphorylation a peut-être une influence sur la régulation ou l'activité des protéines FLIM1, et ainsi souligne éventuellement les différentes fonctions que pourraient avoir les protéines des sous-groupes FLIM1 et WLIM1.

Les protéines du groupe  $\alpha$ LIM1 ont un domaine C-terminal acide. Les domaines LIM1 et LIM2 sont suffisants pour la fixation de la protéine NtLIM1 de tabac sur l'ADN, mais l'activation de la transcription nécessite la protéine entière (Kawaoka *et al.*, 2000). Il a alors été suggéré que le domaine C-terminal acide de cette protéine NtLIM1 puisse avoir une fonction dans l'activation de la transcription de la même manière que les domaines acides des facteurs de transcription VP1 chez le maïs et GAL4 chez la levure (Gill & Ptashne, 1987 ; Ma *et al.*, 1988 ; McCarty *et al.*, 1991). Le domaine C-terminal particulièrement acide des protéines du groupe  $\delta$ LIM2 pourrait avoir aussi la même fonction dans l'activation de la transcription. De façon intéressante, le domaine C-terminal des protéines WLIM2 est différent de celui des autres protéines LIM et est plutôt basique. De même, chez le tabac, la protéine NtWLIM2 serait éventuellement capable de se fixer à l'ADN. En effet, l'ADNc *NtWLIM2* a été isolé avec le promoteur H1 à partir d'un crible d'expression d'ADNc (Mundel *et al.*, 2000), et des expériences de gel retard utilisant le promoteur H4 confirmeraient la fixation de la protéine NtWLIM2 à l'ADN (Gatti *et al.*, 2005). Dans l'hypothèse où les protéines WLIM2 agirait également en tant que facteurs de transcription, la caractéristique de leurs domaines C-terminal avec un pI basique montre qu'elles ont probablement des propriétés particulières dans l'activation de la transcription des gènes.

## 4 Conclusions

Grâce à la découverte de six nouveaux transcrits *PtaLIM* chez *Populus tremula* × *P. alba*, j'ai pu identifier douze modèles de gènes *PtLIM* chez *Populus trichocarpa* qui apparaissent tous dupliqués. Les génomes d'*Oryza sativa* et d'*Arabidopsis thaliana* contiennent respectivement six modèles de gène *LIM*. Par cette analyse bioinformatique, les séquences génomiques des gènes *LIM* chez ces trois espèces ont été actualisées et améliorées. Il est remarquable de constater une forte conservation de la structure génomique de ces gènes avec notamment quatre introns aux positions très précises et des régions codantes 2 et 4 de longueur quasiment identique pour tous les gènes *LIM* analysés. La collection d'ESTs et d'ADNc *LIM* de plantes, puis la détermination des relations phylogénétiques entre les différentes protéines à domaine LIM, fournissent à la communauté scientifique une base de donnée « LIM » utile pour les futures recherches sur les gènes de cette famille. J'ai ainsi défini quatre groupes phylogénétiques ( $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2 et  $\delta$ LIM2) et  $\beta$ LIM1 est un nouveau groupe dont la fonction des gènes reste à explorer. Au sein de ces groupes, nous retrouvons les sous-groupes PLIM1, PLIM2, WLIM1 et WLIM2 précédemment définis, ainsi que de nouveaux sous-groupes tels que FLIM1 et  $\delta$ LIM2 des astéridées, pouvant aussi avoir des spécificités d'expression tissulaire. De plus, il existe des sous-groupes phylogénétiques spécifiques de classes ou sous-classes de plante comme le sous-groupe PLIM1 qui semble être caractéristique des *Asteraceae*. Les protéines LIM issues de monocotylédones et eudicotylédones sont également phylogénétiquement séparées. L'abondance des séquences protéiques collectées nous a permis d'affiner la séquence consensus des domaines LIM, et de nombreux motifs potentiellement intéressants dans la régulation et la fonction de ces protéines ont été identifiés. Cette étude bioinformatique et phylogénétique de la famille de protéines à domaine LIM de plante établit les bases pour les futures analyses fonctionnelles de ces protéines. Or l'analyse de l'expression de ces gènes chez le peuplier est manquante. En effet, il est nécessaire de vérifier l'induction des gènes *GLIMI* dans le bois de tension, de connaître l'expression des gènes  $\beta$ LIM1 et aussi de tester l'expression des gènes  $\delta$ LIM2 dans le pollen. Enfin, une expression préférentielle dans les tissus vasculaires de certains gènes *LIM* est également suspectée.

## **Partie 2. Approche de l'étude fonctionnelle de la famille LIM chez le peuplier par une analyse d'expression de gènes**

### **1 Matériel et méthodes**

#### *1.1. Matériel*

##### 1.1.1. Echantillonnage des organes végétatifs

###### 1.1.1.1. Racine principale et tronc d'un arbre âgé de quatre ans

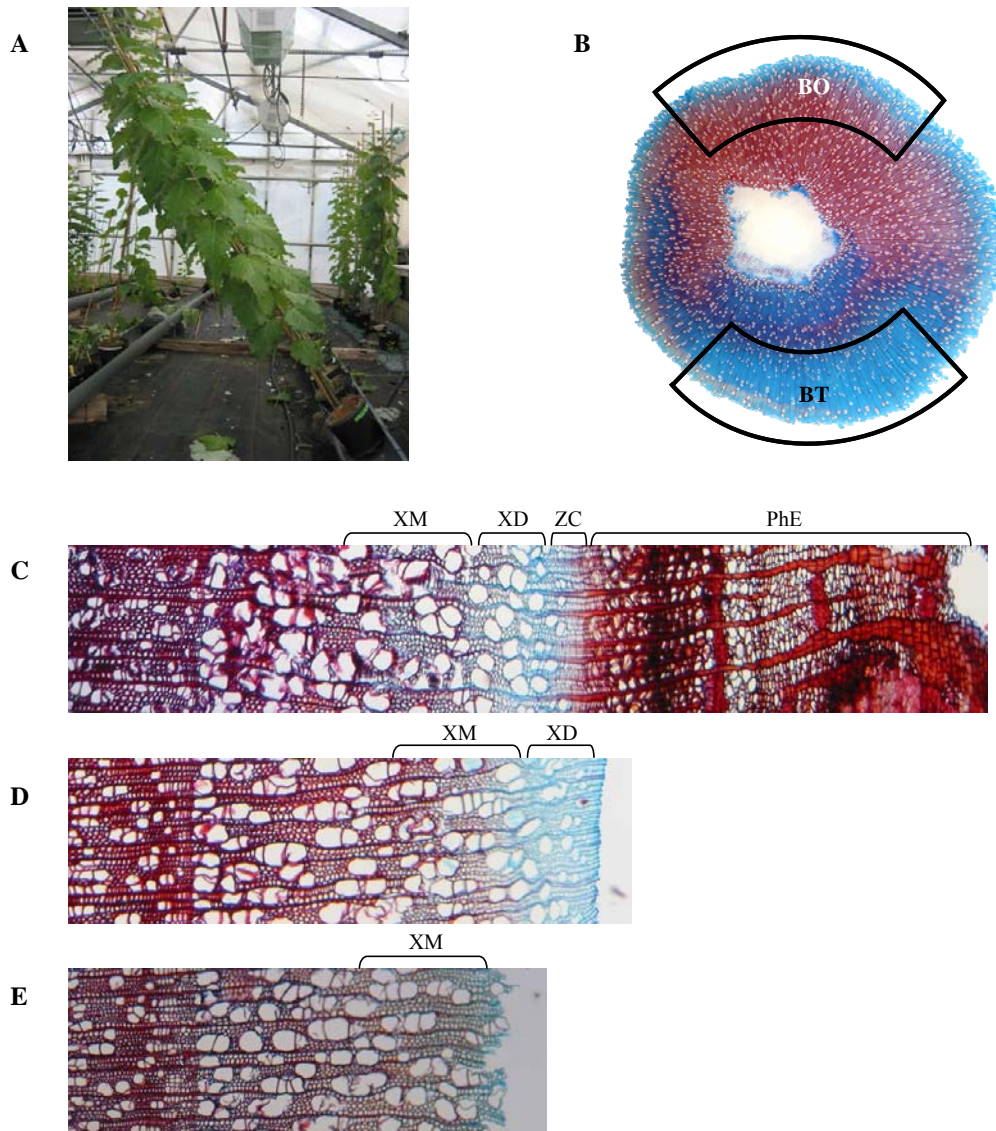
A la fin du mois de mai 2004, pendant la période de croissance active de l'arbre, différents tissus ont été prélevés sur un peuplier grisard, hybride naturel entre peuplier tremble et peuplier blanc (*Populus tremula* × *P. alba*, clone INRA 717-1-B4, clone femelle) cultivé en pépinière et âgé de quatre ans. Le tronc d'un diamètre de 15 cm environ a été écorcé au niveau du cambium pour récupérer les tissus vasculaires secondaires. Après écorçage du tronc, le xylème en développement (XD) composé principalement de cellules en expansion et formant de la paroi primaire et secondaire, a été légèrement raclé à l'aide d'un scalpel. Le xylème mature (XM) dans les tissus plus internes du tronc a été prélevé grâce à un ciseau à bois et contient des cellules à paroi secondaire lignifiée. Les cellules de la zone cambiale (ZC) ont été prélevées en grattant légèrement la face interne de l'écorce. Les échantillons de phloème (Ph) sont facilement séparés de l'écorce (E) par épluchage. D'après les observations histologiques, l'écorçage du tronc s'est effectué au niveau de la zone cambiale. Ainsi, l'échantillon de ZC contiendrait probablement aussi des cellules de phloème et des cellules de xylème en début de différenciation, pendant la phase de division/expansion cellulaire.

Sur le même arbre, nous avons suivi la même méthode de prélèvement pour l'échantillonnage des tissus de la racine principale. La racine principale a été écorcée au niveau du cambium, et le xylème secondaire au centre de la racine a ainsi été séparé de l'écorce racinaire en périphérie contenant le phloème secondaire. Le xylème en développement (XD) a été raclé au scalpel sur le pourtour de la racine écorcée, et le reste de la racine écorcée et raclée constitue le xylème mature (XM) contenant la moelle. A la face



## Partie 2. Analyse d'expression

interne de l'écorce racinaire, la zone cambiale (ZC) a été légèrement raclée au scalpel. L'épiderme n'ayant pas pu être séparé facilement du cortex de la racine, le reste du cortex (PhE) est constitué d'une variété de tissus incluant le phloème secondaire, le phelloderme, et l'épiderme subérisée (ou périderme). Tout ces tissus prélevés sur le tronc et la racine principale ont immédiatement été congelés dans l'azote liquide, et conservés à -80°C jusqu'à utilisation.



**Figure 46.** Dispositif expérimental d'inclinaison des peupliers et méthode d'échantillonnage des tissus du xylème sur des tiges de peupliers inclinés ou droits.

(A) Les pousses de peupliers âgées de deux mois, clone INRA 717-1-B4 (*Populus tremula* × *P. alba*) sont inclinées à 45° par rapport à la verticale par inclinaison des pots, et les tiges sont maintenues par un tuteur pendant un mois. Les arbres témoins (en arrière plan) poussent verticalement.

(B à E) Coupes transversales de jeunes tiges inclinées de peuplier (*P. tremula* × *P. alba*) colorées à la safranine bleu-astra (Photos Laurans F., INRA Orléans). (B) Les zones de prélèvement de bois de tension (BT) et de bois opposé (BO) sont encadrées sur la tige écorcée. (C à E) Grossissement de la face inférieure des tiges correspondant au bois opposé. (C) Tige avant écorçage. (D) Tige écorcée. (E) Tige après prélèvement au scalpel du jeune xylème en développement. PhE : phloème et écorce ; ZC : zone cambiale ; XD : xylème en développement ; XM : xylème mature.

## Partie 2. Analyse d'expression

### 1.1.1.2. Culture *in vitro* : Echantillonnage de cals cellulaires et de jeunes racines de *in vitro* plants

Afin d'étudier des tissus indifférenciés, des cals cellulaires (clone INRA 717-1-B4) cultivés sur milieu MB5 solide (Gamborg *et al.*, 1968) à l'obscurité et à 24°C, ont été prélevés deux semaines après repiquage. Le système racinaire de six plantules (clone INRA 717-1-B4) cultivées *in vitro* a été prélevé et constitue l'échantillon de jeunes racines entières (R) à croissance primaire. Ces plantules ont été cultivées en chambre de culture à 24°C avec une photopériode de 16h de lumière et 8 h d'obscurité, sur du milieu solide MS dilué deux fois (MS/2) et modifié (Murashige & Skoog, 1962 ; Leplé *et al.*, 1992). Ces deux échantillons ont été congelés dans l'azote liquide et conservés à -80°C jusqu'à utilisation.

### 1.1.2. Protocole de mise en tension des arbres et prélèvements des bois tendu et opposé

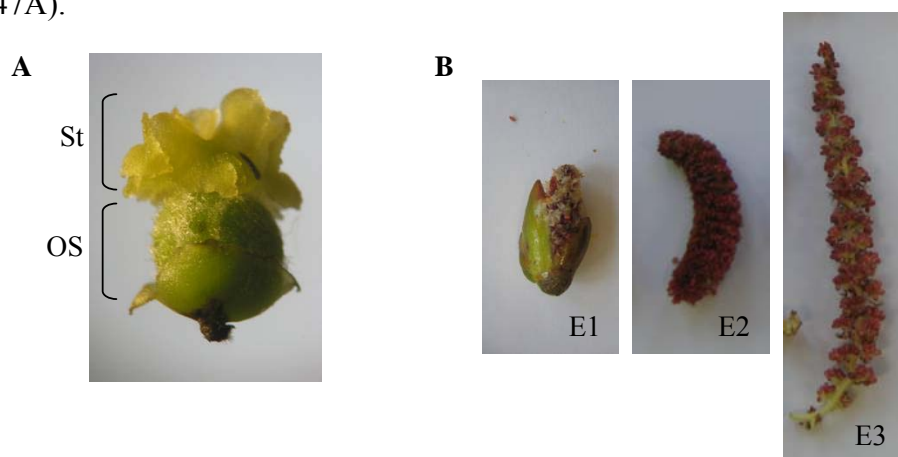
Des plants du clone 717-1-B4 recépés à l'automne 2005 et hivernés à 4°C, ont été transférés en serre le 15 mars 2006, repotés dans des pots de 3 litres et arrosés individuellement avec de l'eau et une substance nutritive « Akaphos<sup>®</sup> Ultra » (COMPO<sup>®</sup>) par un système de goutte à goutte. Le 22 mai, après deux mois et demi de croissance, les pots ont été artificiellement inclinés à 45° par rapport à la verticale et les tiges ont été maintenues par des tuteurs rigides. Un mois après, pour obtenir trois répétitions biologiques, nous avons récolté trois arbres inclinés et trois arbres témoins non inclinés mesurant entre 1,5 et 2 m (Fig. 46A). Des portions de tiges à croissance secondaire mesurant 40 cm et situées 10 cm au dessus du collet ont été écorcées. Le phloème avec l'écorce, le cambium, les xylèmes en développement et mature (Fig. 46C à E) ont été collectés sur tout le pourtour des tiges droites, et sur les faces supérieures et inférieures des tiges inclinées (Fig. 46B), produisant respectivement du bois tendu (BT) et du bois opposé (BO). Le xylème en développement (XD) a été prélevé en raclant légèrement au scalpel la tige écorcée et le xylème mature (XM) est recueilli par un raclage plus profond de la tige. La zone cambiale (ZC) est retirée en raclant légèrement au scalpel la face interne de l'écorce pelée. L'écorce collée au phloème chez ces jeunes plants, n'a pas pu être séparée de celui-ci. Le restant de l'échantillon constitue donc le phloème et l'écorce (PhE). Sur ces mêmes plants inclinés et droits, les parties jeunes (environ 10 cm) de la tige (JTI et JTD) ont été récoltées. Ces parties jeunes encore vertes et

constituées principalement de tissus à croissance primaire (Prassinos *et al.*, 2005) ont été prélevées sous l'apex des tiges entre le premier et le quatrième entrenœud.

Les feuilles ont également été prélevées. Pour cet organe, les pétioles (P), les nervures principales et secondaires ou vaisseaux (V) et le limbe (L) contenant toujours de petites nervures secondaires, ont été séparés au scalpel. Pour tous ces tissus de la tige et des feuilles, les échantillons de chaque arbre ont été congelés et broyés séparément dans l'azote liquide, puis conservés à -80°C. Pour chaque tissu, les échantillons broyés de trois arbres indépendants inclinés ou droits ont été mélangés pour la préparation d'ARN totaux.

### 1.1.3. Echantillonnage des organes reproducteurs

Des branches portant des bourgeons floraux encore fermés ont été prélevées début Avril 2006 sur des clones adultes femelle ou mâle de *Populus nigra* poussant en bord de Loire. Les branches femelles ont été transférées dans du terreau pour l'enracinement des tiges, et permettre le développement des fleurs, leur pollinisation artificielle et le développement ultérieur des graines. Après l'ouverture complète des bourgeons (anthèse) à la fin du mois d'Avril, les fleurs (ou chatons) femelles entières (FF) ont été prélevées avant fécondation. Le rachis, le pédoncule, et la bractée ont été retirés au scalpel. Les stigmates (St) ont été séparés du carpelle contenant le style et l'ovaire (OS) par dissection à l'aide d'une pince et d'un scalpel (Fig. 47A).



**Figure 47.** Echantillonnage des organes reproducteurs femelle et mâle  
(A) Pour les fleurs femelles, le stigmate (St) à été séparé de l'ovaire et du style (OS) où le péricarpe à la base du carpelle n'a pas pu être retiré.

(B) Les fleurs mâles ont été prélevées à différents stades de développement : E1 représente un stade précoce où les chatons sont petits, partiellement ouverts et les étamines sont sessiles. E2 est un stade intermédiaire avec des chatons plus allongés et des étamines encore sessiles sur le rachis. E3 correspond à un stade tardif de développement juste avant l'anthèse avec des chatons pendants et très allongés, portant des étamines étalées sur le rachis.

## Partie 2. Analyse d'expression

Les branches mâles portant des bourgeons floraux fermés ont été transférées dans de l'eau pour permettre la conservation et le développement des étamines jusqu'à l'anthèse correspondant à la libération du pollen. Les fleurs mâles entières (chaton) à différents stades de développement ont été prélevées (Fig. 47B) et les écailles protectrices des fleurs ont été enlevées. Le stade 1 correspond à un stade précoce de développement avec des bourgeons floraux petits partiellement ouverts et portant des étamines peu développées et sessiles sur le rachis (E1). Le stade 2 est un stade intermédiaire où les chatons sont entièrement ouverts, plus allongés, et portent des étamines encore sessiles sur le rachis (E2). Enfin, le stade 3 correspond au stade tardif de développement des étamines, juste avant l'anthèse (E3). A ce stade, les chatons pendants et très longs portent des rachis et des pédoncules allongés avec des étamines plus étalées. Le pollen mature (P) issu de peuplier *P. nigra* var. *italica* a été récolté en 2005 en agitant les étamines matures présentant des anthères ouvertes pouvant facilement libérer le pollen. Pour la germination du pollen et l'obtention de tubes polliniques, les grains de pollen sont réhydratés sous atmosphère humide pendant une heure, puis saupoudrés sur milieu solide MS/2 modifié (Murashige & Skoog, 1962 ; Leplé *et al.*, 1992). Le pollen germé (PG) est récolté au bout de 24 heures de germination en raclant légèrement la surface de l'agar avec une lame de scalpel. La présence de tubes polliniques a été vérifiée sous loupe binoculaire et ceux-ci présentent des tailles variables. A ce stade, environ 80 % des grains de pollen ont germés (Borel C. et Villar M., communication personnelle).

Un mois après la fécondation manuelle des fleurs femelles, les capsules vertes contenant et protégeant les graines sont ouvertes à la main. Puis, les fibres ou cheveux cotonneux (F), sont séparées des graines (G) et des capsules (Ca) à l'aide d'une pince. Ces onze échantillons (fleurs femelles entières, stigmates, carpelles, étamines 1, 2 et 3, pollen mature, pollen germé, capsules, fibres, et graines) ont été immédiatement placés dans de l'azote liquide pour conservation ultérieure à - 80°C.

### 1.2. Méthodes

#### 1.2.1. Analyse de la distribution des ESTs codant pour des protéines à domaine LIM chez le peuplier

J'ai utilisé les ESTs *LIM* récupérées lors de la caractérisation des séquences d'ADNc et de gènes *PtLIM* de peuplier (Voir Matériel et méthode de la partie 1), pour analyser leur

distribution dans les différentes banques d'ESTs. Les clones reséquencés des banques d'ESTs PopulusDB n'ont pas été pris en compte dans cette analyse. Les trois bases de données d'ESTs comportant le plus grand nombre d'ESTs *LIM* et provenant de l'Université British Columbia (Ralph *et al.*, 2006), du centre « Umeå Plant Science Centre » (PopulusDB) (Sterky *et al.*, 2004), et de l'INRA d'Orléans (Déjardin *et al.*, 2004) ont été séparées sur le tableau 14. Sur ce tableau a été reporté les différentes banques d'ESTs indiquant l'origine tissulaire des ESTs. Les ESTs provenant d'autres banques d'ESTs plus minoritaires ont été regroupées dans la liste « Autres banques ».

### 1.2.2. Extraction des ARN totaux de peuplier

Les tissus ont été broyés et homogénéisés dans l'azote liquide en utilisant un mortier et un pilon, et l'ARN total a été préparé en utilisant une méthode adaptée de (Chang *et al.*, 1993). Toutes les étapes ont été réalisées avec des solutions traitées au diéthyl pyrocarbonate (DEPC) à 0,1 %. 200 mg d'extraits sont solubilisés dans 900 µl de tampon d'extraction [Tris-HCl 100 mM à pH 8 ; EDTA 25 mM ; NaCl 2 M ; spermidine 0.5 g/L ; CTAB 2% ; PVP 2% ; β-mercaptoéthanol 2 %] et incubés 30 min à 65°C. Les protéines sont ensuite dénaturées et extraites par l'ajout d'un volume de chloroforme/alcool isoamylique (24/1). Après 5 min de centrifugation à 9100 g, la phase aqueuse prélevée est extraite une seconde fois à l'aide d'un volume de chloroforme/alcool isoamylique (24/1). Les ARN sont ensuite précipités sur la nuit à 4°C par l'addition d'un volume de chlorure de lithium 2 M. Après 20 min de centrifugation à 9100 g à 4°C, le culot d'ARNs est repris dans 100 µl d'eau DEPC. La purification de ces ARN se fait ensuite à l'aide du RNeasy® plant Mini Kit (Qiagen) en suivant les spécifications du constructeur et en utilisant la partie « RNeasy Mini Protocol for RNA Cleanup » du kit.

### 1.2.3. Quantification des ARN totaux

La quantité d'ARN totaux est analysée par spectrophotométrie en mesurant la densité optique à 260 nm avec un NanoDrop® ND-1000 (Nixor Biotech). Une unité de DO à cette longueur d'onde correspond à 40 µg/ml. Les densités optiques à 280 et 230 nm sont aussi mesurées pour évaluer la qualité des ARN en vérifiant la présence de contaminants protéiques absorbant à 280nm ou phénoliques absorbant à 230 nm. La quantité et la qualité des ARN

## Partie 2. Analyse d'expression

totaux extraits sont finalement vérifiées après migration de 5 µg d'ARN sur gel d'agarose non dénaturant à 1,2 % dans du tampon TBE [Tris-Borate 45 mM ; EDTA 1 mM].

### 1.2.4. Reverse transcription des ARNm en ADNc

Deux microgrammes d'ARN totaux isolés sont reverse transcrits en ADNc en utilisant 500 ng d'amorces oligo(dT)<sub>12-18</sub>, des dNTP à 0,5 mM, du « First Strand Buffer » 1X (Invitrogen™), du DTT à 10 mM, et 200 unités de SuperScript™ II RT RNase H<sup>-</sup> dans un volume total de 20 µl et en suivant les recommandations du fabricant (Invitrogen™). Le mélange réactionnel est incubé pendant 1 h à 42°C et la réaction est stoppée 15 min à 70°C. Après la synthèse de l'ADNc simple brin, l'ARN matrice complémentaire de l'ADNc est dégradé en incubant la réaction à 37°C pendant 20 min avec 2 unités de *E. coli* RNase H (Invitrogen™).

### 1.2.5. Caractéristiques du choix des amorces pour les RT-PCR

Chaque couple d'amorces a été conçu pour amplifier spécifiquement un gène *PtLIM*. Les amorces ont été sélectionnées manuellement et à l'aide du logiciel Primer3 sur le site [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Leur température d'hybridation a été calculée avec le programme Oligonucleotide Properties Calculator (<http://www.basic.northwestern.edu/biotools/oligocalc.html>). Ces deux logiciels ont permis de tester les interactions possibles entre amorces. Celles-ci ont été choisies en fonction des paramètres suivants : Le contenu en base GC des amorces doit être compris entre 40 et 60 % (avec un optimum choisi à 50 %), et un fort contenu en GC dans le tiers de la séquence en 5' a été opté. Pour une forte spécificité d'hybridation, la taille des amorces a été choisie entre 22 et 24 pb, et elles contiennent au moins deux GC dans les cinq dernière bases en 3'. Enfin, les amorces possèdent une température d'hybridation (T<sub>m</sub>) élevée proche de 60°C, avec une variation de T<sub>m</sub> entre couple d'amorces ne dépassant pas 4°C. Les couples d'amorces ont été choisis de telle manière que le fragment d'ADN amplifié ait une taille comprise entre 300 et 600 pb. Pour faciliter les expérimentations, les tailles d'amplification et les températures d'hybridation sont les mêmes pour tout les gènes *PtLIM*, à l'exception des gènes *PtβLIM1a* et *b* et *PtβLIM2a* et *b*. Toutes les amorces ont été conçues de telle manière qu'elles puissent amplifier des ADNc provenant à la fois de peuplier hybride *Populus tremula* × *P. alba* et de

## Partie 2. Analyse d'expression

l'espèce *P. trichocarpa*. Ainsi, on a de fortes chances d'amplifier les matrices RT issues d'échantillons de l'espèce *P. nigra*. Les amorces ont été positionnées aux extrémités de la séquence codante (CDS) des gènes de tel manière à amplifier les quatre introns dans le cas d'une contamination par de l'ADN génomique. Pour les gènes fortement similaires entre eux tel que les gènes dupliqués, certaines amorces sont situées dans les régions 3' ou 5' UTR plus variables. Pour chaque gène *PtLIM* dupliqué, une amorce commune en 5' et une amorce spécifique en 3' (ou *vice versa*) ont été choisies. Enfin, pour vérifier la spécificité des amorces, celles-ci ont été soumises à une recherche BlastN contre la séquence du génome de *P. trichocarpa*. Les amorces ont été synthétisées par l'entreprise OPERON™ Biotechnologies.

**Tableau 13.** Séquences des oligonucléotides utilisés pour les analyses de RT-PCR semi quantitative.

Gène	Taille prédite du produit PCR (bp)	Température d'hybridation des amorces appliquée	Nombre de cycles PCR	Amorces sens et antisens (de 5' à 3')
<i>PtWLIM1a</i>	634	65	25	GACAAGTGTTCGGCTGCAAGAAC et AGGCAGGAAAATGCAGGATGAACA
<i>PtWLIM1b</i>	625	65	30	GACAAGTGTTCGGCTGCAAGAAC et GCAAGGCAGGAAAATACACTCGTT
<i>PtGLIM1a</i>	590	65	30	CCTGTTGATAATGAGAACGCCAGC et CAGCAATCTCTACAGCAGTGATGC
<i>PtGLIM1b</i>	583	65	30	CCTGTTGATAATGAGAACGCAAGC et CACGGCGGTGATATTCACAGTG
<i>PtβLIM1a</i>	444	60	35	TTAGATGCCACCACTGCAAG et CTGGTGTAACATGTTTCGTGCT
<i>PtβLIM1b</i>	436	60	35	TTAGATGCCACCACTGCAAG et AGATGTTTCATGTTTGCCGAA
<i>PtWLIM2a</i>	646	65	30	GGTGATCCGAGACAGGTAAACAGT et GAGATGTTTAAGCCTCTGGCACAG
<i>PtWLIM2b</i>	639	65	40	TCCAAAGGAGTTAACAGGACAGTC et GAGATGTTTAAGCCTCTGGCACAG
<i>PtPLIM2a</i>	609	65	35	CATTCACGGGAACCCTGGATAAG et TGCTCTTCTGACTTTTCCTCCTCC
<i>PtPLIM2b</i>	562	65	40	CATTCACGGGAACCCTGGATAAG et TCGGCCAGTTCAGGAGGTGG
<i>PtβLIM2a</i>	434	60	35	GGACCTTAGAGAAATGCAAGG et GATACCATCAAGAGCTGCATAA
<i>PtβLIM2b</i>	426	60	35	GGACCTTAGAGAAATGCAAGG et CAAGAGCAGCGTAGGATGAT
<i>PopFLA1</i>	923	55	20	GTTTCTAGCATTATATCTTAACAAC et CAGAATTATTCAACACACAATCGGC
<i>18SRNA</i>	232	55	20	CTTCGGGATCGGAGTAATGA et GCGGAGTCCTAGAAGCAACA

### 1.2.6. Test des amorces sur vecteur et mise au point des conditions PCR

Pour les douze gènes *PtLIM*, l'amplification spécifique de chaque couple d'amorces a été testée sur les ADNc disponibles (*PtaGLIM1a* et *b*, *PtaWLIM1a* et *b*, *PtaWLIM2a* et *PtaPLIM2a*) insérés dans le vecteur pTriplex. Pour chaque gène, différentes températures d'hybridation (60, 65 et 68°C), ainsi que différents temps d'hybridation des amorces (30, 45 ou 60 sec), et la concentration en MgCl<sub>2</sub> (1, 1,25 et 1,5 mM) ont été testées. Les conditions optimales pour une amplification spécifique ont été obtenues pour une concentration en MgCl<sub>2</sub> de 1,25 mM et une durée d'hybridation de 45 sec. Les températures d'hybridation pour chaque gène sont reportées dans le tableau 13. Les conditions de saturation de l'amplification PCR ont été ajustées en faisant varier le nombre de cycles PCR.

### 1.2.7. Réaction de polymérisation en chaîne « Polymerase chain reaction » (PCR)

Chaque réaction d'amplification PCR a été réalisée avec une matrice d'ADNc simple brin dilué au 1/60<sup>e</sup>, 0,5 unités de rTaq DNA polymerase (Invitrogen™), du MgCl<sub>2</sub> à 1,25 mM, des dNTP à 0,2 mM, et des amorces gène spécifique à 0,2 µM (Tableau 13), dans un volume total de 20 µl. Pour tous les gènes, la réaction PCR débute par une étape initiale de dénaturation des ADN de 5 min à 94°C. Les conditions de cycle PCR varient en fonction des gènes *PtLIM* (Tableau 13). Les ADNc *PtLIM* sont amplifiés par 25 à 40 cycles d'amplification comprenant une étape de dénaturation de 45 sec à 94°C, une étape d'hybridation des amorces à 65°C ou 60°C pendant 45 sec, et une étape de synthèse de l'ADN à 72°C pendant 1 min. L'ARNr *18S* et le gène *PopFLAI* sont respectivement utilisés comme témoin de la quantité de matrice RT et comme gène marqueur de la formation de bois de tension (Lafarguette *et al.*, 2004). L'ARNr *18S* est amplifié par 20 cycles constitués de 45 sec à 94°C, 1 min à 55°C, et 1 min 30 à 72°C. Le gène *PopFLAI* est amplifié par 20 cycles constitués de 45 sec à 94°C, 1 min 30 à 55°C, et 1 min 30 à 72°C (Lafarguette *et al.*, 2004). La réaction PCR se termine par une étape d'extension des doubles brins d'ADN à 72°C pendant 10 min et est finalement maintenue à 4°C. L'amplification a été réalisée en utilisant un appareil GeneAmp® PCR System 9700 (Applied Biosystems). 10 µl de chaque produit PCR sont mis à migrer par électrophorèse sur un gel d'agarose à 1 % dans du tampon TBE.



## Partie 2. Analyse d'expression

Après incubation du gel dans une solution de bromure d'éthidium à 50  $\mu\text{g}/\mu\text{l}$  pendant 15 min, le gel est lavé dans de l'eau distillée pendant 20 min, puis visualisé sous lampe UV.

### 1.2.8. Séquençage des produits RT-PCR

Pour chaque gène *PtLIM*, les produits RT-PCR obtenus à partir de matrices de xylème en développement ou d'étamines matures, ont été séquencés afin de confirmer l'identité des gènes amplifiés. Le séquençage a été réalisé en 5' ou 3' à partir de produits RT-PCR purifiés, en utilisant les amorces spécifiques de chaque gène et à l'aide du kit BigDye<sup>®</sup> terminator (Applied Biosystems) en suivant les instructions du fabricant. Les produits de séquençage obtenus sont purifiés puis analysés avec un séquenceur automatique à capillaires ABI PRISM<sup>®</sup> 3100 DNA 16-capillary Genetic Analyser (Applied Biosystems). La migration des fragments par électrophorèse capillaire passe par une cellule de détection qui analyse la fluorescence émise par les fluorophores différents pour chaque désoxyribonucléotide. Un chromatogramme est généré ainsi qu'un fichier texte.

### 1.2.9. Etude de l'expression par Northern Blot du gène *PtGLIM1a*

#### 1.2.9.1. Migration des ARN en conditions dénaturantes et transfert des ARN sur membrane

Dix microgrammes d'ARN totaux sont séchés et repris dans 100 µl de tampon de charge [MOPS 1X (MOPS 20 mM, acétate de sodium 5 mM, EDTA 1 mM) formamide 50%, formaldéhyde 4,5%, glycérol 7%, bleu de bromophénol 0,5%, bromure d'éthidium 0,025%], puis dénaturés pendant 10 min à 70°C. Les ARN sont séparés par électrophorèse sur gel d'agarose dénaturant [agarose 1,2 %, MOPS 1X, formaldéhyde 3%]. La migration se fait à 70 V dans du tampon MOPS 1X. Après rinçage du gel dans de l'eau DEPC, les ARN sont transférés par capillarité dans du NaOH à 50 mM, sur une membrane Hybond-N+ (Amersham Biosciences<sup>®</sup>).

#### 1.2.9.2. Préparation des sondes radiomarquées

La sonde correspondant au gène *PopFLA5* a été d'abord amplifiée par PCR à partir du clone d'ADNc PtaXM0020G6G0614 (CF229065) à l'aide des amorces 5'-GAAAAGCCAACAAAGGCAGT-3' et 5'-TGGAATCTCAAAACTCAACTCAA-3'. Vingt-

## Partie 2. Analyse d'expression

cinq nanogrammes de produit PCR ont été marqués avec du [<sup>32</sup>P]-dCTP (50 µCi) par « random priming » avec le « Kit Prime-a-gene<sup>®</sup> » (Promega<sup>®</sup>) selon les instructions du fabricant. La sonde correspondant au gène *PtGLIM1a* a été marquée directement par « random priming » à partir d'environ 200 ng de clone plasmidique PtaJXT0019G11G11 (EF035035) avec du [<sup>32</sup>P]-dATP (50 µCi) et en utilisant les kits « Kit Strip-EZ<sup>™</sup> DNA, Random Primed StripAble<sup>™</sup> DNA Probe Synthesis and Removal kit » (Ambion) selon les instructions du fabricant.

### 1.2.9.3. Hybridation de la membrane

La membrane est préhybridée dans un four à hybridation pendant 2h à 65°C dans du tampon [Na<sub>2</sub>HPO<sub>4</sub> 0,5 M à pH7, BSA 1%, SDS 7%, EDTA 1mM, ADN de sperme de saumon dénaturé 67 µg/ml]. La sonde radioactive, préalablement dénaturée pendant 5 min à 95°C, est ajoutée au tampon précédent pour l'hybridation de la membrane à 65°C toute la nuit. La membrane est ensuite lavée 3 fois 15 min, successivement dans du tampon SSC 2X à pH 7, SDS 0,1%, puis du SSC 1X à pH 7, SDS 0,1%, et enfin du SSC 0,1X à pH 7, SDS 0,1%. La membrane est exposée à un film autoradiographique à -80°C pendant 12h pour la sonde *PopFLA5* et 7 jours pour la sonde *PtGLIM1a*. La membrane, d'abord hybridée avec *PopFLA5*, a été décapée avec du SDS 0,1% porté à ébullition, avant hybridation avec *PtGLIM1a*.

## 2 Résultats

### 2.1. Analyse de la distribution des ESTs codant pour des protéines à domaine LIM chez le peuplier

La première étape pour étudier la fonction des gènes codant pour des protéines à domaine LIM est d'analyser dans quels tissus ils sont exprimés. En effet, le profil d'expression d'un gène nous permet de connaître les processus développementaux et/ou physiologiques pour lesquels ce gène pourrait être éventuellement impliqué. Avec le nombre croissant d'ESTs de peuplier disponibles dans les bases de données publiques (plus de 350000 ESTs en avril 2007), il est actuellement possible d'avoir une vue globale de l'abondance des transcrits *LIM* dans les différentes banques d'ESTs.

Partie 2. Analyse d'expression

**Tableau 14.** Répartition des ESTs de peuplier codant pour des protéines à domaine LIM dans les différentes banques d'ESTs présentes dans les bases de données publiques (Déjardin *et al.*, 2004; Ralph *et al.*, 2006; Sterky *et al.*, 2004).

Banques/Gènes	PtWLIM1a	PtWLIM1b	PtGLIM1a	PtGLIM1b	PtβLIM1a	PtβLIM1b	PtWLIM2a	PtWLIM2b	PtPLIM2a
Zone cambiale (ZC)	1	2	0	1	0	0	0	0	0
Xylème en développement de bois opposé (XD-BO)	1	0	1	1	0	0	0	0	0
Xylème en développement de bois tendu (XD-BT)	4	5	7	4	0	0	1	0	1
Xylème mature (XM)	0	0	7	2	0	0	0	0	0
Total	6	7	15	8	0	0	1	0	1
PT-GT-FL-A-3 et PTxD-IL-N-A-9	0	0	3	0	0	0	1	2	0
PT-DX-A-7 et PT-DX-N-A-10	2	1	7	1	0	1	1	2	0
PT-P-FL-A-2 et PTxN-IB-N-A-11	8	4	9	5	0	0	2	3	0
PTxD-NR-A-8	0	0	0	1	0	0	0	1	0
PTxD-ICC-N-A-14	1	0	0	0	5	0	0	3	0
PT-MB-N-A-15	0	0	0	0	1	2	1	0	0
Total	11	5	19	7	6	3	5	11	0
Zone cambiale (A + B) et cambium actif (UB)	0	3	0	10	0	0	1	3	3
Bois de tension (G)	0	2	11	3	0	0	2	1	0
Cellules mortes du bois (X)	0	0	9	0	0	0	0	0	0
Racines (R)	0	1	1	0	0	0	0	0	0
Chatons femelles (M) et méristème des tiges (T)	0	0	0	0	0	0	0	2	0
Feuilles stressées par le froid	1	0	0	0	0	0	0	0	0
Total	1	6	21	13	0	0	3	6	3
Autres banques	14	2	3	1	0	2	3	1	1
Total des banques	<b>32</b>	<b>20</b>	<b>58</b>	<b>29</b>	<b>6</b>	<b>5</b>	<b>12</b>	<b>18</b>	<b>5</b>

La nomenclature des protéines à domaine LIM de peuplier est issue de l'analyse phylogénétique (Voir Partie 1). Les banques d'ESTs de l'INRA d'Orléans (Déjardin *et al.*, 2004), de l'Université British Columbia (UBC) (Ralph *et al.*, 2006) et du centre de recherche « Umeå Plant Science Centre » (Sterky *et al.*, 2004) sont respectivement représentée en haut, milieu et bas de tableau. La liste des banques de l'UBC est la suivante : **PT-GT-FL-A-3**: Feuilles jeunes et matures, avec des bouts de tige vert, **PTxD-IL-N-A-9**: Feuilles matures locales récoltées après alimentation continue par *M. disstria*, **PT-DX-A-7** et **PT-DX-N-A-10**: Xylème extérieur récolté deux fois par semaine entre avril et octobre 2002, **PT-P-FL-A-2**: Phloème et cambium, **PTxN-IB-N-A-11**: Ecorce (avec le phloème et le cambium attaché) récoltée après alimentation continue par *C. lapathi*, **PTxD-NR-A-8**: Racines récoltées à partir d'arbres âgés de trois mois cultivés en milieu hydroponique dépourvu de source d'azote pendant 24 et 48 h, ainsi que des arbres cultivés sur milieu normal, **PTxD-ICC-N-A-14**: Cellules en culture poussant sur milieu supplémenté avec de l'acide salicylique, du benzothiadazole, du méthyl jasmonate, du chitosan ou des extraits de *Pollacia radiosa*, **PT-MB-N-A-15**: Bourgeons végétatifs terminaux. Pour les détails des origines des banques provenant de l'UBC, se référer à (Ralph *et al.*, 2006).

## Partie 2. Analyse d'expression

Ces banques d'ESTs ont été préparées à partir de différents tissus/organes de peupliers comme par exemple les racines, les feuilles, et les différentes parties du tronc incluant le phloème, le cambium, et le xylème (Tableau 14). Nous avons donc analysé dans les différentes banques d'ESTs de peuplier le nombre d'ESTs correspondant aux différents gènes codant pour des protéines à domaine LIM. Par défaut, les banques d'ADNc ne sont pas normalisées ni enrichies, et peuvent donc être représentative de la population de messenger dans l'échantillon. Comme les clones d'ADNc ont été tirés au hasard puis séquencés en 5' ou 3', cette distribution d'ESTs est le reflet du niveau d'expression des gènes dans les différents tissus incluant le bois. Jusqu'à Juillet 2006, j'ai identifié au total 186 ESTs de peuplier codant pour des protéines à domaine LIM. Certaines banques contiennent peu d'ESTs *LIM* comme celles de « racines » (deux ESTs) ou de « feuilles stressées par le froid » (une EST). Par contre, d'autres banques contiennent un nombre plus élevé d'ESTs *LIM* comme la banque de « xylème en développement de bois tendu » (22 ESTs) et les banques PT-P-FL-A-2 et PTxN-IB-N-A-11 construites à partir de cambium, de phloème et d'écorce (31 ESTs). J'ai respectivement trouvé six, sept, et neuf transcrits correspondant à différents gènes *LIM* de peuplier, dans nos séquences d'ESTs publiées (Déjardin *et al.*, 2004), dans la base de données d'ESTs PopulusDB (<http://www.populus.db.umu.se>) (Sterky *et al.*, 2004), et dans d'autres bases de données incluant notamment des ESTs provenant de l'Université British Columbia (UBC) (Ralph *et al.*, 2006). Il est important de préciser qu'aucune EST correspondant aux gènes *PtPLIM2b*, et *PtβLIM2a* et *b* n'a été trouvée dans les banques d'ESTs de peuplier. Les banques d'ESTs étant principalement réalisées à partir de tissus prélevés sur le tronc, ceci indiquerait que ces gènes sont probablement peu ou pas exprimés dans le phloème ou le xylème.

Le résultat le plus marquant est que *PtGLIM1a* est le gène pour lequel j'ai identifié le plus d'ESTs (58), comparé aux autres gènes *PtLIM* (Tableau 14). Dans nos banques d'ESTs, parmi les consensus codant des protéines à domaine LIM, *PtaGLIM1a* est le plus gros consensus avec quatorze ESTs. Ces ESTs proviennent majoritairement de banques issues de xylème en développement de bois tendu et de xylème mature (avec sept ESTs chacune), des tissus prélevés sur des arbres âgés de deux ans et inclinés pendant huit mois. Le profil est le même dans la base de données PopulusDB avec neuf ESTs provenant d'une banque de cellules mortes du bois, et onze ESTs issues d'une banque de bois tendu contenant surtout des cellules formant de la paroi secondaire et une couche G. Dans la base de données de l'UBC, les ESTs *PtGLIM1a* proviennent majoritairement de xylème, avec sept ESTs, et de cambium/phloème avec neuf ESTs. Ce gène est également exprimé dans une moindre mesure

dans d'autres tissus tels que les racines, les feuilles et le bois opposé. Son dupliqua, le gène *PtGLIM1b* semble aussi être fortement exprimé avec un total de 29 ESTs. Ces ESTs sont aussi retrouvées dans nos banques de bois tendu ainsi que dans celle de la base de données PopulusDB avec quatre et trois ESTs respectivement. Avec dix ESTs *PtGLIM1b* retrouvées dans la banque issue de zone cambiale de PopulusDB, et cinq ESTs dans les banques de cambium/phloème de l'UBC, ce gène semble être préférentiellement exprimé dans la zone cambiale, contrairement au gène *PtGLIM1a* dont les ESTs sont plutôt trouvées dans des banques de xylème.

Globalement, les deux autres consensus *PtaWLIM1a* et *PtaWLIM1b* présents dans nos banques contiennent une majorité d'ESTs issues de xylème tendu, et aucune EST ne provient de xylème mature. Les ESTs *PtWLIM1a*, sont fortement représentées dans les banques de phloème/cambium construites à l'UBC (huit ESTs). Et étonnamment, un seul singleton, correspondant au gène *PtWLIM1a* et issu de la banque « feuilles stressées par le froid », a été isolé dans les banques d'ESTs PopulusDB contenant pourtant un grand nombre d'ESTs. Les autres banques d'ESTs contiennent quatorze ESTs *PtWLIM1a* provenant principalement de phloème, de xylème et de tiges. Le gène dupliqué *PtWLIM1b*, avec un total de 20 ESTs, semble être moins exprimé dans le peuplier que le gène *PtWLIM1a* (32 ESTs). Parmi les cinq ESTs *PtWLIM1b* identifiées dans la base de données d'ESTs de l'UBC, quatre ESTs proviennent des banques de cambium/phloème. En analysant nos banques d'ESTs et celles de PopulusDB, on peut voir que ce gène est exprimé dans la zone cambiale (cinq ESTs) et le bois de tension (sept ESTs).

Pour les gènes *PtβLIM1a* et *b*, un faible nombre d'ESTs a été séquencé, principalement à partir des banques construites à l'UBC. Il est intéressant de voir que le gène *PtβLIM1a* n'est pas représenté par des ESTs issues de tissus formant du bois, mais par des ESTs presque exclusivement issues de cellules en culture (cinq ESTs sur six). Des tissus variés, comme le xylème, les bourgeons végétatifs et les racines expriment le gène *PtβLIM1b*.

Dans nos banques d'ESTs, *PtaWLIM2a* est un singleton issu de la banque de xylème en développement de bois tendu. La banque bois de tension de PopulusDB contient aussi deux ESTs *PtaWLIM2a*. Ce gène est par ailleurs aussi exprimé dans d'autres parties de l'arbre comme les feuilles, le cambium, les bourgeons végétatifs. Les ESTs correspondant au gène *PtWLIM2b* ne sont pas trouvées dans nos banques. Dans les autres banques d'ESTs, elles proviennent de différents tissus tels que le xylème, le cambium, le phloème et les bourgeons végétatifs des tiges, les chatons femelles, les feuilles, et les cultures cellulaires. Ces deux

gènes, semblent donc exprimés constitutivement chez le peuplier et plutôt faiblement, par comparaison aux gènes *PtGLIM1a* et *b* et *PtWLIM1a* et *b*.

Dans notre banque d'ESTs, le gène *PtPLIM2a* est représenté par un singleton issu de la banque de xylème en développement de bois tendu. Des ESTs correspondant au gène *PtPLIM2a* ont également été trouvées dans les racines de peuplier (Kohler *et al.*, 2003) et dans les banques de zone cambiale de la base de données PopulusDB (trois ESTs). Ce gène semble donc moins exprimé dans le bois que les autres gènes LIM de peuplier.

### 2.2. Etude de l'expression des gènes LIM de peuplier par RT-PCR semi quantitative

L'analyse de la distribution des ESTs LIM de peuplier dans les différentes banques d'ESTs publiques montre que les gènes *PtWLIM1a* et *b* et *PtGLIM1a* et *b* sont exprimés dans les tissus vasculaires du tronc (phloème, cambium et xylème), alors que les gènes *PtβLIM1a* et *b* sont plutôt exprimés dans les cultures cellulaires ou les bourgeons végétatifs, et les gènes *PtWLIM2a* et *b* sont plutôt exprimés de manière ubiquitaire. Toutefois, l'analyse de la distribution des ESTs dans les différents tissus de peuplier est limitée et imprécise. En effet, les banques ont été préparées de façon indépendante par différents organismes, sur des espèces et des génotypes de peuplier différents et avec des protocoles expérimentaux différents. Ainsi, cette méthode ne nous donne que des indications sur l'expression des gènes LIM de peuplier, et la comparaison de l'expression d'un gène entre différents tissus est difficile. Enfin, certains transcrits comme *PtPLIM2a* et *b* et *PtβLIM2a* et *b* sont peu ou pas représentés dans les différentes banques d'ESTs. C'est pourquoi, une étude plus précise de l'expression des gènes *PtLIM* dans les différents tissus de peuplier a été entreprise en utilisant la technique de RT-PCR semi quantitative.

#### 2.2.1. Analyse des profils d'expression des gènes *PtLIM* dans les tissus végétatifs de peupliers.

Au niveau des tissus vasculaires de la tige de jeunes peupliers, le phloème et l'écorce (PhE), la zone cambiale (ZC), le xylème en développement (XD) et le xylème mature (XM) ont été prélevés sur de jeunes peupliers cultivés en serre et âgés de trois mois (Fig. 48). Par comparaison, des prélèvements similaires ont été effectués sur un peuplier âgé de quatre ans ayant un tronc bien développé sur lequel il a été facile de séparer l'écorce (E) du phloème

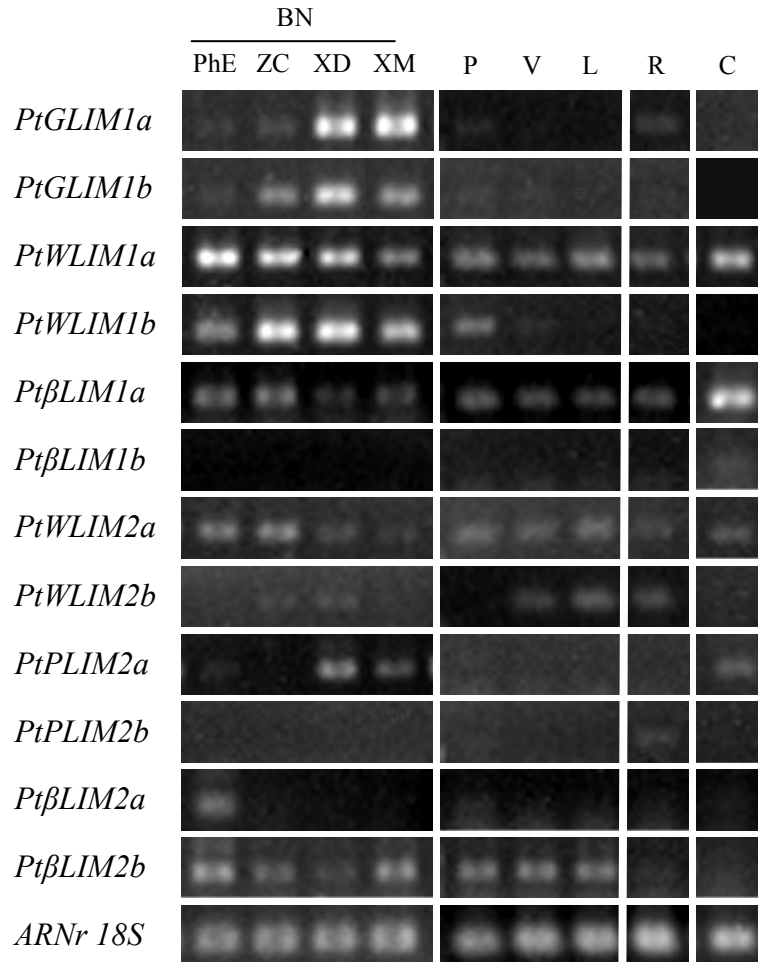
(Ph) (Fig. 49). D'autre part, sur ces mêmes peupliers cultivés en serre, les pétioles (P), les nervures ou vaisseaux (V) et le limbe (L) des feuilles ont été analysés séparément afin de mesurer la quantité des transcrits *PtLIM* dans les tissus vasculaires de la feuille. Les racines étant peu développées et de petite taille sur ces plants cultivés en pot, il était difficile d'en isoler les différents tissus (xylème, cambium et phloème) pour les analyses d'expression. Des jeunes racines entières (R) ont alors été prélevées sur des *vitro* plants ayant poussé sur milieu gélosé afin d'éviter d'éventuelles contaminations par des microorganismes du sol. De plus, sur la racine pivot ou principale de l'arbre âgé de quatre ans, il a été possible d'isoler les différents tissus vasculaires à croissance secondaire : le xylème mature (XM) incluant la moelle, le xylème en développement (XD), la zone cambiale (ZC), et « l'écorce racinaire » constituée du phloème secondaire et de l'épiderme (PhE). Enfin, des cals cellulaires (C) ont été récoltés afin d'étudier l'expression des gènes *PtLIM* dans un tissu indifférencié. L'expression des gènes a été analysée par RT-PCR semi quantitative. Les conditions d'amplification PCR ont été mises au point pour chaque gène *PtLIM* (Cf. Matériel et méthodes). Chaque amorce amplifie spécifiquement son gène cible. Ceci a été vérifié d'une part en testant l'amplification PCR de chaque couple d'amorces sur les plasmides portant les différents ADNc *PtaLIM*, et d'autre part en réalisant pour chaque gène un séquençage des produits RT-PCR obtenus à partir d'ARN totaux de peupliers (résultats non montrés). L'amplification des ADNc de l'ARNr 18S et du gène *PopFLAI* (Lafarguette *et al.*, 2004) ont été utilisés pour contrôler la quantité des produits RT de chaque échantillon.

L'analyse de l'expression des gènes *PtLIM* par RT-PCR semi quantitative nous montre que, dans la tige ou le tronc de peuplier, le gène *PtGLIM1a* est fortement exprimé dans le xylème en développement et le xylème mature quel que soit l'âge du peuplier (Fig. 48 et 49). Par contre son expression est plus faible dans le phloème et le cambium voire nulle dans l'écorce indiquant une expression du gène plus spécifique des étapes tardives de la formation du xylème secondaire. Au niveau des feuilles, le gène *PtGLIM1a* n'est que très faiblement exprimé dans les pétioles et son expression est légèrement plus élevée dans les jeunes racines de peuplier cultivés *in vitro*. Au contraire, au niveau d'une racine plus âgée formant des tissus vasculaires secondaires, le gène *PtGLIM1a* est fortement exprimé dans le xylème en développement (Fig. 49). Son dupliqua, le gène *PtGLIM1b* est aussi fortement exprimé dans le xylème en développement de la tige, du tronc et de la racine, mais il est plus faiblement exprimé dans le xylème mature contrairement au gène *PtGLIM1a*. Il est également exprimé dans le phloème et le cambium. La quantité d'ARNm *PtGLIM1b* est par ailleurs très faible dans les pétioles, nervures et jeunes racines. On peut toutefois dire que ces deux gènes



## Partie 2. Analyse d'expression

sont aussi exprimés, à un niveau très faible, dans les tissus vasculaires des feuilles. L'amplification des deux transcrits est mieux détectée sur des pétioles et nervures d'un peuplier adulte cultivé en pépinière (résultats non montrés). Ainsi, dans les organes végétatifs, ces deux gènes dupliqués *PtGLIM1a* et *b* semblent préférentiellement exprimés dans les tissus vasculaires à croissance secondaire.



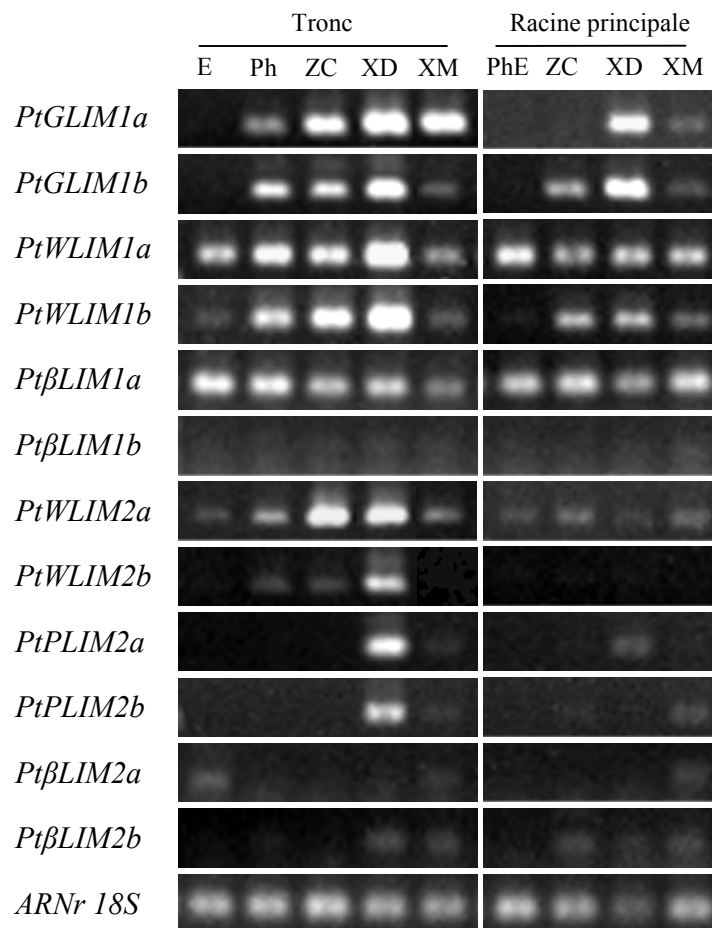
**Figure 48.** Analyse de l'expression des gènes *PtLIM* par RT-PCR semi quantitative dans les organes végétatifs de jeunes peupliers cultivés en serre.

Sur les tiges de trois peupliers droits (*Populus tremula* × *P. alba*) produisant du bois « normal » (BN) et cultivé en serre pendant trois mois, le phloème et l'écorce (PhE), la zone cambiale (ZC), le xylème en développement (XD) et le xylème mature (XM) ont été prélevés. Dans les feuilles, les tissus vasculaires comme les pétioles (P), les vaisseaux (V), ont été séparés du limbe (L). L'analyse d'expression inclut des jeunes racines entières (R) issues de six peupliers cultivés *in vitro* et des cals cellulaires (C) cultivés à l'obscurité. L'expression de l'*ARNr 18S* est montrée comme témoins de la quantité des ARN totaux.

Chez de jeunes peupliers cultivés en serre, le gène *PtWLIM1a* est plus exprimé dans le phloème et le cambium que dans le xylème. Son expression diminue progressivement en fonction de la différenciation du xylème et est relativement faible dans le xylème mature. Par contre son expression est forte dans le xylème en développement d'un arbre adulte montrant que l'expression du gène peut varier en fonction de l'âge de l'arbre. Il présente une expression

## Partie 2. Analyse d'expression

constitutive avec la présence de transcrits *PtWLIM1a* dans d'autres tissus tel que l'écorce, les différentes parties de la feuille et de la racine, mais la quantité de transcrit reste plus faible dans les feuilles que dans les tiges. Enfin, une quantité importante de transcrit est détectée dans les cals cellulaires dépourvus de tissus vasculaires. Le gène *PtWLIM1b* dupliqué a un profil d'expression différent de celui du gène *PtWLIM1a*. *PtWLIM1b* est fortement exprimé dans le cambium et le xylème en développement et son expression diminue graduellement vers le phloème jusqu'à l'écorce et vers le xylème mature. Ce gène semble donc préférentiellement exprimé lors des étapes précoces de différenciation vasculaire vers le phloème ou le xylème. De plus, un profil d'expression similaire est observé dans les tissus vasculaires de la racine. Dans les feuilles, on ne détecte la présence d'ARNm *PtWLIM1b* que dans les pétioles et aussi très faiblement dans les nervures.



**Figure 49.** Analyse de l'expression des gènes *PtLIM* par RT-PCR semi quantitative dans les organes végétatifs d'un arbre âgé de quatre ans.

Sur le tronc d'un peuplier non incliné (*Populus tremula* × *P. alba*) cultivé en pépinière pendant quatre ans, le phloème (Ph) a été séparé de l'écorce (E), et la zone cambiale (ZC), le xylème en développement (XD) et le xylème mature (XM) ont été prélevés au scalpel. Au niveau de la racine pivot, le xylème secondaire en développement (XD) a été récolté au scalpel et le reste de la racine écorcée constitue le xylème mature (XM). La zone cambiale (ZC) a été prélevée au scalpel sur la face interne de l'écorce racinaire, constituée de phloème secondaire et de l'épiderme (PhE). L'expression de l'*ARNr 18S* est montrée comme témoin de la quantité des ARN totaux.

## Partie 2. Analyse d'expression

L'expression du gène *PtβLIM1a* dans les tissus vasculaires n'est pas aussi forte que celle des gènes précédemment cités. En effet, le nombre de cycle PCR utilisé pour ce gène est plus grand que ceux utilisés pour les gènes *PtGLIM1a* et *b* et *PtWLIM1a* et *b*. La quantité d'ARNm *PtβLIM1a* est toutefois plus importante dans le phloème, l'écorce et dans une moindre mesure la zone cambiale que dans le xylème. Dans les feuilles et les jeunes racines, la quantité de transcrit est la même quelque soient les tissus analysés. Le gène *PtβLIM1a* est aussi fortement exprimé dans les tissus vasculaires secondaires de la racine et dans les tissus indifférenciés comme les cals cellulaires. A l'opposé, le gène dupliqué *PtβLIM1b* n'est pas du tout exprimé dans la tige de peuplier, ni dans les feuilles ou les racines. Une très faible expression est toutefois visible dans le xylème mature de la racine et les cals cellulaires.

Au niveau des tiges de peuplier, les gènes *PtWLIM2a* et *PtWLIM2b* ont des profils d'expression similaires à ceux des gènes *PtWLIM1a* et *PtWLIM1b* respectivement, mais avec un niveau d'expression bien plus faible. Dans les tiges de jeunes peupliers cultivés en serre, *PtWLIM2a* est plus exprimé dans le phloème et le cambium que dans le xylème, tandis que chez un arbre plus âgé l'expression du gène est plus forte dans le cambium et le xylème en développement que dans le phloème, l'écorce et le xylème mature. Concernant le gène *PtWLIM2b*, on détecte un faible signal dans le cambium et le xylème en développement. Comme pour le gène *PtWLIM2a*, l'amplification observée des ADNc *PtWLIM2b* dans le xylème en développement est plus forte chez un arbre cultivé en pépinière. Dans les feuilles, les racines, et les cals, l'expression du gène *PtWLIM2a* semble être constitutive comme celle de *PtWLIM1a*. Par contre, *PtWLIM2b* est plutôt exprimé dans le limbe que dans les tissus vasculaires (nervures et pétiole) de la feuille. On détecte aussi son expression dans les jeunes racines, mais pas dans les différents tissus de la racine plus développée.

Le gène *PtPLIM2a* est exprimé dans le xylème et la quantité de transcrit est plus importante dans le xylème en développement que dans le xylème mature. Le xylème en développement de la racine ainsi que les cals cellulaires expriment aussi le gène *PtPLIM2a*. Son expression est nulle dans les feuilles et les jeunes racines. Son dupliqua, le gène *PtPLIM2b* ne semble pas du tout exprimé dans la tige et les feuilles de peupliers cultivés en serre. Par contre, de façon surprenante, il est bien exprimé dans le xylème en développement et plus faiblement dans le xylème mature chez un arbre plus âgé cultivé en pépinière. On peut voir une très faible quantité de transcrit dans les racines jeunes et dans le xylème mature de racines plus âgées.

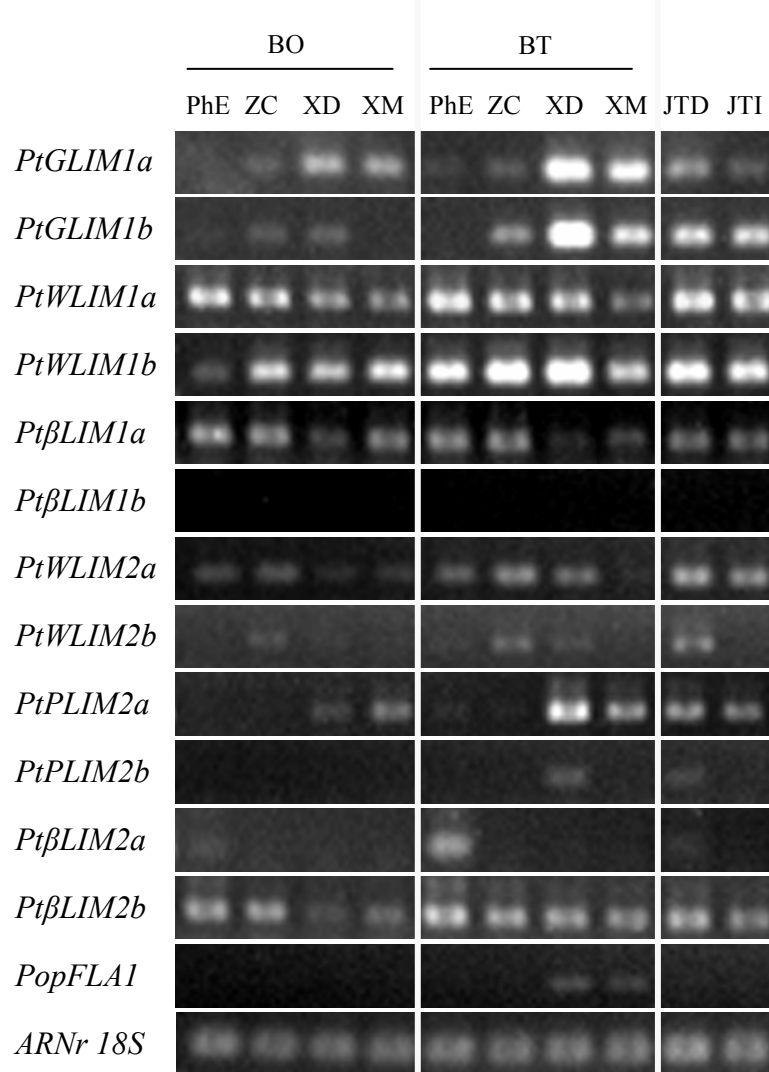
Les gènes *PtβLIM2a* et *b* ont un profil d'expression particulier. Au niveau des tiges, le gène *PtβLIM2a* est faiblement exprimé dans le phloème et l'écorce. L'analyse de l'expression

du gène au niveau du tronc nous montre que l'expression du gène est plus spécifique de l'écorce, et par ailleurs, un très faible signal est observé dans le xylème mature du tronc et de la racine principale. La quantité de transcrits *PtβLIM2b* est plus abondante dans le phloème/écorce et le xylème mature que dans le cambium et le jeune xylème des jeunes tiges. L'expression du gène *PtβLIM2b* serait préférentielle des étapes tardives de la différenciation du xylème ou du phloème. Étonnamment, au niveau du tronc d'un peuplier adulte, le gène est beaucoup plus faiblement exprimé et uniquement dans le xylème en développement et mature. Mis à part un très faible signal dans les pétioles, *PtβLIM2a* ne semble pas exprimé dans les feuilles. *PtβLIM2b* est lui exprimé de manière constitutive dans les feuilles et à un niveau très faible dans les jeunes racines. Une faible expression est également observée dans la racine plus âgée, de la zone cambiale jusqu'au xylème mature. Enfin, ces deux gènes ne sont pas exprimés dans les calcs cellulaires.

### 2.2.2. Analyse de la quantité de transcrits *PtLIM* dans les bois tendu et opposé prélevés sur de jeunes arbres

Une précédente étude de l'équipe a montré que les ESTs similaires au gène codant une protéine spécifique du pollen SF3 et correspondant au gène *PtGLIM1a*, sont plus abondantes dans le xylème en développement de bois de tension (Déjardin *et al.*, 2004). D'autre part, il a été montré récemment que le gène *PttGLIM1b* est également plus exprimé dans le bois de tension (Andersson-Gunneras *et al.*, 2006). Dans le but d'analyser finement le profil d'expression des gènes *PtLIM*, notamment *PtGLIM1a* et *b*, au cours des étapes de la formation du bois de tension, nous avons réalisé des prélèvements de phloème incluant l'écorce, de zone cambiale, de xylème en développement, et de xylème mature. Ces prélèvements ont été réalisés sur des arbres inclinés pendant un mois, à la fois sur la face supérieure de la tige où se forme le bois de tension et sur la face inférieure où l'on trouve le bois opposé dépourvu de fibres G. L'expression des gènes *PtLIM* a été analysée par RT-PCR semi quantitative. Le profil d'expression du gène *PopFLA1* est également figuré. Ce gène est exclusivement exprimé dans le xylème en développement et mature de bois tendu (Fig. 50). On peut donc considérer ce gène *PopFLA1* comme un marqueur de la formation de bois de tension chez le peuplier (Lafarguette *et al.*, 2004), et il est utilisé ici comme témoin montrant que du bois de tension a bien été formé à la face supérieure des tiges inclinées.

## Partie 2. Analyse d'expression



**Figure 50.** Analyse de l'expression des gènes *PtLIM* par RT-PCR semi quantitative dans les bois opposé et tendu.

Sur les tiges de trois arbres (*Populus tremula* × *P. alba*) cultivés en serre pendant un mois et inclinés pendant 15 jours, le phloème et l'écorce (PhE), la zone cambiale (ZC), le xylème en développement (XD) et le xylème mature (XM) ont été prélevés sur la face inférieure de la tige produisant du bois opposé (BO) et sur la face supérieure produisant du bois tendu (BT). L'expression des gènes *PtLIM* au niveau de la croissance primaire de la tige a été analysée sur des portions jeunes, encore vertes et situées sous l'apex, de trois tiges inclinées (JTI) ou droites (JTD). L'expression de l'*ARNr 18S* est montrée comme témoin de la quantité des ARN totaux et l'expression du gène *PopFLAI* est utilisé comme témoin de l'induction de l'expression des gènes dans le bois de tension (Lafarguette *et al.*, 2004).

Tout d'abord, nos résultats confirment la forte induction du gène *PtGLIM1a* dans le xylème en développement de bois tendu. Mais aussi, *PtGLIM1a* est fortement induit dans le bois de tension au niveau du xylème mature. Enfin, bien que l'expression du gène *PtGLIM1a* soit plus forte dans les parties jeunes des tiges d'arbres droits que dans celles d'arbres inclinés, son niveau d'expression est nettement plus faible dans les zones de croissance primaire de la tige que dans les zones de croissance secondaire. Le gène dupliqué *PtGLIM1b* montre un profil d'expression similaire à *PtGLIM1a* avec une forte induction de l'expression du gène dans le xylème en développement et le xylème mature de bois tendu. Par contre, dans

le bois opposé, l'amplification du gène *PtGLIM1b* n'est observée que dans le cambium et le xylème en développement, et ce gène n'est pas exprimé dans le xylème mature de bois opposé. Dans les parties jeunes de la tige, la quantité de transcrite est relativement forte et ne varie pas en fonction de l'inclinaison de l'arbre.

Concernant le gène *PtWLIM1a*, le profil d'expression du gène dans les différents tissus, coté bois opposé ou tendu, est similaire au profil observé dans le bois d'arbre droit avec une expression du gène qui diminue du phloème jusqu'au xylème mature. Cependant, l'expression du gène *PtWLIM1a* semble légèrement induite dans le phloème/écorce, le cambium et le xylème en développement de la face supérieure de la tige, tandis que dans le xylème mature le niveau d'expression du gène ne semble pas varier entre le bois de tension et le bois opposé. L'expression du gène *PtWLIM1b* est très fortement induite dans la partie supérieure des tiges inclinées au niveau du phloème, de la zone cambiale, et du xylème tendu en développement. De plus, l'expression du gène semble même être réprimée dans le xylème en développement de bois opposé. Par contre, l'expression du gène *PtWLIM1b* ne paraît pas être modifiée dans le xylème mature de bois tendu, par rapport à celui de bois opposé. Pour les deux gènes *PtWLIM1a* et *b*, on observe aussi une forte accumulation des transcrits dans les tissus jeunes de la tige en croissance primaire comparable à celle observée dans les tissus à croissance secondaire.

La mise en inclinaison des arbres ne semble pas modifier significativement le profil d'expression du gène *PtβLIM1a*. En effet, que ce soit à la face supérieure ou inférieure de la tige inclinée, le profil d'expression du gène est relativement similaire à celui observé pour un arbre droit avec une plus forte expression dans le phloème/écorce et le cambium que dans le xylème. Cependant, l'expression du gène *PtβLIM1a* semble diminuer sensiblement dans les xylèmes en développement et mature de bois tendu. Le gène *PtβLIM1a* est aussi exprimé dans les parties jeunes des tiges d'arbres inclinés ou droits. Comme nous l'avons vu pour les tiges d'arbres droits, le gène dupliqué *PtβLIM1b* n'est également pas exprimé dans les tissus de la tige d'arbres inclinés.

Comme dans le cas des tiges droites, sur la face inférieure des tiges inclinées, le gène *PtWLIM2a* est légèrement plus exprimé dans le phloème/écorce et dans la zone cambiale que dans le xylème. Par contre, à la face supérieure des tiges inclinées, l'expression du gène *PtWLIM2a* est induite dans la zone cambiale et le xylème en développement. De même, le gène *PtWLIM2a* est significativement exprimé dans les parties jeunes des tiges droites et inclinées. Par contre, son expression est plus faible dans le xylème mature et quasiment indétectable dans le xylème mature de bois tendu. Le gène *PtWLIM2b* est faiblement exprimé

dans le cambium et le xylème en développement de bois tendu tandis que son expression semble plus faible dans le xylème en développement de bois opposé. *PtWLIM2b* est exprimé dans les parties jeunes des tiges droites, mais ne l'est pas dans celles des tiges inclinées.

Dans le bois de tension, l'expression du gène *PtPLIM2a* est fortement induite dans le xylème en développement, et l'induction de l'expression du gène est plus faible dans le xylème mature. La quantité de transcrite dans le xylème en développement de bois opposé est réduite par rapport à celle de bois normal. Le gène est également exprimé dans les parties jeunes de la tige. Pour le gène *PtPLIM2b*, on détecte une faible accumulation de transcrite dans le xylème tendu en développement, alors que dans le bois d'arbre droit, le gène ne semble pas exprimé. Une faible amplification est aussi visible dans les jeunes parties de la tige d'arbres non inclinés.

Le gène *PtβLIM2a*, exprimé dans le phloème/écorce de la tige, présente une quantité plus importante de transcrite dans la partie supérieure de la tige inclinée. Par ailleurs, il semble très faiblement exprimé dans les parties jeunes de la tige d'arbres droits. Au niveau du phloème et du xylème mature, l'expression du gène *PtβLIM2b* est légèrement plus forte dans le bois de tension que dans le bois opposé. Par contre, son expression est nettement plus forte dans le xylème en développement de bois tendu que dans celui de bois opposé. Enfin, *PtβLIM2b* est exprimé dans les parties jeunes de la tige, un peu plus fortement dans celles d'arbres droits que dans celles d'arbres inclinés.

### 2.2.3. Analyse de la quantité de transcrits *PtLIM* dans les organes reproducteurs du peuplier

Des études d'expression ont montré que chez les plantes, certains gènes *LIM* ont une spécificité d'expression dans le pollen. C'est le cas par exemple de *HaPLIM1a* (aussi nommé SF3) chez le tournesol (Baltz *et al.*, 1992a ; Baltz *et al.*, 1999) et PGPS/D1 chez le pétunia (Guyon *et al.*, 2000). Les gènes *LIM* ont ainsi été classés en quatre sous groupes : PLIM1 et PLIM2 spécifiquement exprimés dans le pollen et WLIM1 et WLIM2 exprimés dans toute la plante (Eliasson *et al.*, 2000). L'étude phylogénétique des gènes *LIM* chez les plantes nous montre que chez le peuplier, il n'existe pas de gènes classés dans le sous groupe PLIM1, et qu'il existe deux gènes classés dans le sous groupe PLIM2 (*PtPLIM2a* et *b*). D'autre part, deux autres gènes classés dans le groupe βLIM2 (*PtβLIM2a* et *b*) sont phylogénétiquement proches du sous groupe PLIM2 et pourraient eux aussi être exprimés dans le pollen. Par ailleurs, d'autres études ont montré que chez *Arabidopsis* et le colza (*Brassica napus*, une autre

Brassicassée), les gènes *AtWLIM2a* et *BnWLIMI* sont fortement exprimés dans les graines (Dean Rider *et al.*, 2003 ; Huang *et al.*, 2007). Enfin, chez le cotonnier, un grand nombre d'EST homologues à différents gènes *LIM* proviennent de banques d'ADNc de fibres de coton à différents stades de développement. L'analyse de l'expression des gènes *LIM* de peuplier dans les organes reproducteurs a été réalisée sur des prélèvements effectués à partir d'individus adultes mâles et femelles de *P. nigra* var. *italica* et *P. nigra*. En effet, le peuplier est une espèce dioïque comportant des individus soit mâle soit femelle et nous n'avions pas à notre disposition de clone mâle 717-1-B4 (*P. tremula* × *P. alba*). Les organes reproducteurs mâles analysés sont d'une part des fleurs mâles portant des étamines (portant des anthères) à différents stades de développement, et d'autre part, des grains de pollen non germé, ou germé formant des tubes polliniques. Pour les organes reproducteurs femelles, les fleurs femelles ouvertes ont été disséquées pour séparer les ovaires et les styles des stigmates. Les capsules contenant les graines et les fibres cotonneuses ont également été analysées séparément.

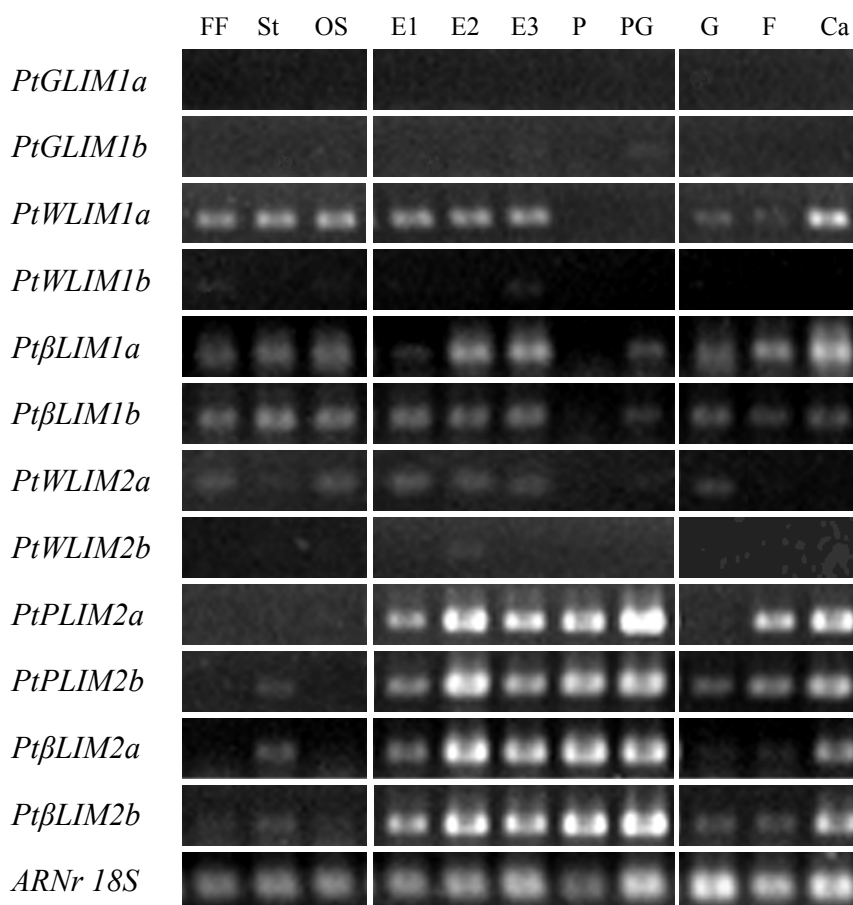
Tout d'abord, les gènes *PtGLIM1a* et *b* ne sont pas exprimés dans les organes reproducteurs, excepté un très léger signal au niveau des grains de pollen germés pour le gène *PtGLIM1b* (Fig. 51). Ces deux gènes semblent bien être spécifiques du bois chez le peuplier. Le gène *PtWLIM1a* est exprimé avec la même intensité dans les fleurs femelles (ovaire/style et stigmate) et dans les étamines quelque soit leurs stades de développement. On observe une faible expression du gène dans les graines et les fibres, tandis qu'elle est plus forte dans la capsule entourant la graine. A l'opposé, l'expression du gène *PtWLIM1b* dans les organes reproducteurs est très faible et est restreinte à l'ovaire/style et les étamines à un stade tardif de développement.

Les gènes *PtβLIM1a* et *b* ont un profil d'expression similaire au niveau des organes mâle et femelle. Tout les deux sont exprimés avec la même intensité au niveau des stigmates et des ovaires/styles. Ils sont également exprimés dans les étamines, mais contrairement à *PtβLIM1b*, l'expression du gène *PtβLIM1a* augmente avec le développement des étamines. On trouve, pour ces deux gènes, une très faible expression dans les grains de pollen formant des tubes polliniques, alors qu'ils ne semblent pas exprimés dans le pollen mature (non germé). Après fécondation, *PtβLIM1a* et *b* sont exprimés dans les graines, les fibres et aussi les capsules. On peut notamment remarquer la forte expression du gène *PtβLIM1a* dans les capsules des graines. Le gène *PtβLIM1b* est donc spécifiquement exprimé dans les organes reproducteurs de peuplier (excepté le pollen non germé) et n'est pas du tout exprimé lors du développement végétatif de l'arbre.



## Partie 2. Analyse d'expression

Les gènes dupliqués *PtWLIM2a* et *b* ne sont pas exprimés de la même manière dans les organes reproducteurs. Le gène *PtWLIM2a* est exprimé dans les fleurs femelles au niveau de l'ovaire et du style, et dans les fleurs mâles, il est exprimé à un niveau constant quelque soit le développement des étamines. On observe aussi une faible expression dans les graines. Par contre, le gène *PtWLIM2b* ne semble pas être exprimé dans les organes reproducteurs. On constate uniquement un faible signal dans les étamines au stade 2.



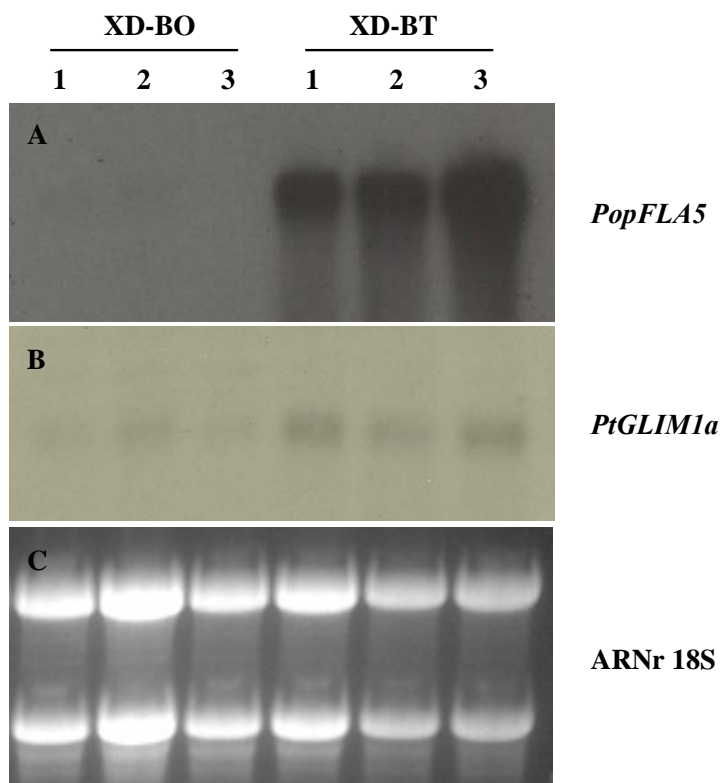
**Figure 51.** Analyse de l'expression des gènes *PtLIM* par RT-PCR semi quantitative dans les organes reproducteurs.

Lors des prélèvements effectués sur des peupliers matures (*P. nigra* var. *italica* et *P. nigra*), les fleurs femelles entières (FF) ont été disséquées afin de séparer l'ovaire et le style (OS) du stigmate (St). Différents stades de développement des inflorescences mâles ont été analysés: des fleurs à un stade précoce, partiellement ouvertes et ayant des étamines sessiles sur le rachis (E1), des fleurs ouvertes, plus allongées avec des étamines encore sessiles sur le rachis (E2), et des fleurs ouvertes, très allongées, à un stade pré-anthèse et avec des étamines étalées sur le rachis (E3) Enfin, du pollen mature non germé (P), et du pollen germé (PG) présentant des tubes polliniques ont également été analysés. Enfin, l'expression des gènes *PtLIM* a été étudiée au niveau des graines (G), des fibres cotonneuses de peuplier (F) et des capsules (Ca) protégeant les graines. L'expression de l'*ARNr 18S* est montrée comme témoin de la quantité des ARN totaux.

Les gènes du groupe  $\beta$ LIM2 (*PtPLIM2a* et *b* et *PtβLIM2a* et *b*) ont un profil d'expression similaire au niveau des organes reproducteurs mâles. Ils ont tous la caractéristique d'être très fortement exprimés dans les étamines à des stades tardifs de différenciation, et surtout ce sont les seuls gènes *LIM* de peuplier très fortement exprimés

## Partie 2. Analyse d'expression

dans le pollen et les tubes polliniques (Fig. 51). Dans les fleurs femelles, on observe une faible expression uniquement des gènes *PtPLIM2b* et *PtβLIM2a* et *b* dans le stigmate. Au niveau des capsules contenant les graines, le profil d'expression de ces gènes est variable. *PtPLIM2b* est exprimé dans les capsules, les fibres de coton, et à un niveau plus faible dans les graines, alors que *PtPLIM2a* est fortement exprimé uniquement dans les fibres et les capsules. Les gènes *PtβLIM2a* et *b* sont faiblement exprimés dans les graines et les fibres et leur expression est forte dans les capsules.



**Figure 52.** Analyse par Northern-Blot de l'expression des gènes *PopFLA5* et *PtGLIM1a* dans le xylème en différenciation de bois opposé et tendu.

Dix microgrammes d'ARN totaux provenant de xylème en différenciation de bois opposé (XD-BO) et de bois tendu (XD-BT) prélevés sur trois tiges (1, 2, 3) de peupliers inclinés ont été séparés par électrophorèse, colorés avec du bromure d'éthidium (C) et transférés sur une membrane nylon (A et B). La membrane a été successivement hybridée avec des sondes spécifiques des gènes *PopFLA5* (A) (Lafarguette et *al.*, 2004) et *PtGLIM1a* (B). La sonde précédente a été décapée de la membrane avant toute nouvelle hybridation. Les membranes ont été révélées par autoradiographie après exposition pendant 12 h (A) et une semaine (B).

### 2.3. L'analyse en Northern blot confirme l'induction de l'expression du gène *PtGLIM1a* dans le bois de tension

L'expression du gène *PtGLIM1a* dans le xylème en développement des bois opposé et tendu a été vérifiée par Northern blot (Fig. 52). Les résultats obtenus sur trois peupliers indépendants confirment les résultats RT-PCR montrant l'induction de l'expression du gène

*PtGLIM1a* dans le xylème en développement de bois tendu (XD-BT). Le faible signal observé au niveau du xylème en développement tant du côté bois opposé que du côté bois tendu est dû à une faible intensité de la radioactivité du dCTP-[<sup>32</sup>P] lors de l'hybridation avec la sonde *PtGLIM1a*. Toutefois, ces résultats RT-PCR et l'analyse de la distribution des ESTs nous montrent que le gène *PtGLIM1a* est plus faiblement exprimé dans le xylème que les gènes *PopFLA1* à *10* (Déjardin *et al.*, 2004 ; Lafarguette *et al.*, 2004). De plus, cette induction dans le xylème tendu est nettement moins forte que celle observée pour le gène *PopFLA5* (Fig. 52).

### 3 Discussion

#### 3.1. L'analyse de la répartition des ESTs révèle la spécificité d'expression vasculaire des gènes LIM chez le peuplier.

Les résultats d'analyses de la distribution des ESTs codant pour des protéines à domaine LIM dans les différentes bibliothèques d'ESTs nous informent sur les profils d'expression de ces gènes dans les différents tissus. Les résultats nous montrent que les gènes *PtWLIM1a* et *b* et *PtGLIM1a* et *b* sont abondamment exprimés dans les tissus formant le xylème ou le phloème (Tableau 14). L'effort de séquençage d'une grande quantité d'ADNc de peuplier par les différentes équipes de recherche a été principalement réalisé à partir de tissus formant du bois. Toutefois, ces études ciblées sur ce tissu révèlent ainsi pour la première fois, une possible implication des gènes LIM dans le développement vasculaire des arbres. D'autre part, avec une grande quantité d'ESTs provenant de bibliothèques issues de bois de tension, le gène *PtGLIM1a* pourrait jouer un rôle dans la formation de ce bois particulier. Les ESTs correspondant aux gènes *PtWLIM2a* et *b* sont retrouvées dans le bois mais aussi dans d'autres tissus tel que les feuilles. Enfin, un faible nombre d'ESTs correspondant aux gènes *PtβLIM1a* et *b* et *PtPLIM2a* a été séquencé. Cependant, la majorité des ESTs *PtβLIM1a* sont issues de tissus indifférenciés tels que les cultures cellulaires (Tableau 14). Alors que l'analyse phylogénétique classe le gène *PtPLIM2a* au sein du sous groupe PLIM2 spécifique du pollen, et que les résultats RT-PCR indiquent une forte expression du gène dans les anthères et le pollen mature, il est surprenant de constater que la bibliothèque « V » de la banque d'ESTs PopulusDB, construite à partir de chatons mâles de *P. trichocarpa* ne contienne aucune EST correspondant à ce gène. Il est possible que cette bibliothèque n'ait été réalisée qu'à partir de jeunes bourgeons mâles n'ayant pas encore des anthères bien développées. De plus,

aucune banque d'ADNc issue de pollen de peuplier n'est actuellement disponible. L'absence d'EST correspondant aux gènes *PtPLIM2b* et *Pt $\delta$ LIM2a* et *b*, peut également être expliquée par le manque de banques d'ADNc de peuplier provenant de pollen. Le manque de représentativité tissulaire des banques d'ADNc est un point important à souligner. En effet, le peuplier et les autres espèces ligneuses sont principalement étudiés pour comprendre la formation du bois, un matériel économiquement important. Par conséquent, la majorité des banques d'ADNc proviennent de tissus vasculaires prélevés sur le tronc et les banques issues d'autres tissus tels que les feuilles ou les organes reproducteurs sont sous représentées. La comparaison de l'abondance des ESTs entre le tissu bois et les autres tissus est donc biaisée. D'autre part, pour les gènes qui ne sont représentés que par un faible nombre d'ESTs, la comparaison du nombre d'ESTs trouvées entre les différents tissus n'est pas envisageable. Toutefois, ces résultats de distribution d'ESTs ainsi que ceux des analyses phylogénétiques nous a donné des informations précieuses sur le type de tissu à analyser en priorité lors des expériences de RT-PCR semi quantitatives.

### *3.2. L'étude de l'expression des gènes *PtLIM* par RT-PCR semi quantitative révèle des spécificités d'expression tissulaire particulières.*

Comme il a été suggéré pour les gènes *LIM* de plantes (Eliasson *et al.*, 2000), les résultats d'étude d'expression par RT-PCR nous montrent que chaque gène *PtLIM* a une propre spécificité d'expression tissulaire. Les gènes *PtGLIM1a* et *b* sont spécifiques du xylème des tiges et sont fortement induit dans le bois de tension (Fig. 50). Leurs proches homologues, les gènes *PtWLIM1a* et *b* sont fortement exprimés dans les tissus vasculaires de la tige. Mais alors que l'expression du gène *PtWLIM1b* semble être restreinte aux tissus vasculaires et est induite dans le bois de tension, le gène *PtWLIM1a* est aussi exprimé de manière ubiquitaire dans d'autres tissus tels que les fleurs ou le limbe (Fig. 48 et 51). Au sein du nouveau groupe phylogénétique  $\beta$ LIM1, l'expression du gène *Pt $\beta$ LIM1b* est limitée aux organes reproducteurs et au contraire, celle de *Pt $\beta$ LIM1a* est plutôt constitutive. Bien qu'il ne soit pas exprimé dans le pollen, les fibres cotonneuses et les capsules des graines, le gène *PtWLIM2a* présente aussi un profil d'expression constitutif. L'expression du gène *PtWLIM2b* est faible et restreinte au cambium et xylème des tiges et aux feuilles. Les gènes *PtPLIM2a* et *b* et *Pt $\delta$ LIM2a* et *b* sont les seuls gènes exprimés dans le pollen mature (Fig. 51). Très fortement exprimé dans les organes reproducteurs mâle, ces gènes sont aussi exprimés dans d'autres tissus comme le xylème pour *PtPLIM2a* (Fig. 48).

### 3.2.1. Corrélations entre les résultats d'analyse de distribution d'EST et les résultats RT-PCR

Dans l'ensemble, les résultats de RT-PCR semi quantitative sont en accord avec l'analyse de la distribution des ESTs *LIM* chez le peuplier. La forte expression des gènes *PtGLIM1a* et *b* et *PtWLIM1a* et *b* dans le cambium et le xylème en développement ou mature des tiges reflète bien l'abondance des ESTs correspondantes dans ces tissus (Fig. 48). Par exemple, lorsque nous regardons les distributions d'ESTs, le gène *PtGLIM1b* semble plus exprimé dans la zone cambiale que le gène *PtGLIM1a*, ce qui apparaît également en analyse RT-PCR. Il en est de même concernant la forte expression du gène *PtGLIM1b* dans le xylème mature. De plus, le caractère constitutif de l'expression du gène *PtWLIM2a*, ainsi que la forte expression du gène *PtβLIM1a* dans les cals cellulaires sont cohérents avec la distribution des ESTs observée pour ces deux gènes. L'étude de l'expression des gènes *LIM* a été réalisée chez le peuplier hybride *Populus tremula* × *alba* et l'espèce *Populus nigra* var. *italica*. Or, il serait faux de penser que les résultats d'expression obtenus ici soient exactement les mêmes pour une autre espèce de peuplier ayant une morphologie et un écotype différents. La régulation de l'expression des gènes *LIM* chez le peuplier, notamment des gènes dupliqués, peut varier entre les différentes sections *Populus* et même au sein d'espèces appartenant à la même section. Ceci est illustré par l'analyse de la répartition des ESTs *PtWLIM1a* qui est différente entre *P. tremula* × *tremuloides* et *P. tremula* × *alba* ou *P. trichocarpa* (Tableau 14). D'autre part, les résultats RT-PCR concernant le gène *PtβLIM1b*, ne reflètent pas la distribution des ESTs correspondantes chez *P. trichocarpa* ou *P. euphratica*. Ceci est probablement dû au faible nombre d'ESTs séquencées et au manque de banques issues d'organes reproducteurs. Enfin, alors que la distribution des ESTs suggère que le gène *PtWLIM2b* est plus exprimé dans le tronc que le gène *PtWLIM2a*, les résultats RT-PCR nous indiquent plutôt le contraire (Fig. 49).

Par ailleurs, on peut noter que la plupart des banques d'ADNc de peuplier ont été construites à partir de prélèvements réalisés sur des arbres adultes âgés d'au moins un an, et au contraire, la plupart de nos prélèvements ont été effectués sur des arbres jeunes de trois mois. Une différence d'âge des arbres pourrait aussi expliquer des différences observées entre la distribution des ESTs et les analyses d'expression de gènes par RT-PCR. De même, on observe de légères différences dans le profil d'expression des gènes *PtLIM* entre un peuplier cultivé en pépinière âgé de quatre ans, et les peupliers cultivés en serre âgés de seulement trois mois. Ainsi, la forte expression des gènes *PtWLIM2a*, et *b* et *PtPLIM2b* dans le xylème

en développement d'un arbre âgé de quatre ans, contraste avec l'expression plus faible observée pour ces gènes dans le même tissu d'arbres plus jeunes. Ce résultat, ainsi que l'absence d'expression du gène *PtδLIM2b* dans le phloème et le cambium d'un arbre âgé, sont difficilement interprétables car beaucoup de paramètres rentrent en jeu. Ces différences d'expression observées peuvent être dues à un effet de l'âge de l'arbre ou bien aux conditions écophysologiques qui sont différentes entre les peupliers cultivés en serre et ceux cultivés en pépinière.

Enfin, le protocole expérimental pour la culture des plantes et l'échantillonnage des différents tissus prélevés sont cruciaux, surtout en regard du tissu bois extrêmement complexe. La façon de racler les différents tissus du cambium ou du xylème peut varier entre les manipulateurs. Les différences d'expression localisée surtout au niveau du cambium et du xylème peuvent aussi être expliquées par une possible variation de l'endroit où se situe l'écorçage. Cette variation peut-être fonction de l'âge de l'arbre, ou bien de l'intensité de l'activité cambiale différente selon les périodes de l'année. Une standardisation stricte des prélèvements ou bien l'application de techniques de prélèvement plus précises comme la réalisation de fines coupes tangentielles s'avèrent donc importantes pour l'étude de la différenciation du xylème ou du phloème (Hellgren *et al.*, 2004).

### 3.2.2. L'expression des gènes *PtLIM* est régulée en fonction des différentes étapes de différenciation du xylème et du phloème

Dans l'ensemble, l'expression des gènes *PtLIM* dans le bois est forte au niveau du xylème en développement, et décroît de part et d'autre du côté de la zone cambiale et du côté du xylème mature. Ceci est surtout vrai pour les gènes *PtGLIM1b*, *PtWLIM1b*, *PtPLIM2a*, *PtWLIM2b*, et dans une moindre mesure pour le gène *PtWLIM2a*. Lorsqu'on regarde dans le détail les résultats obtenus précédemment par une analyse microarray de l'expression des gènes chez le peuplier (*Populus tremula* × *P. tremuloides*) au cours des processus de xylogénèse (Hertzberg *et al.*, 2001), on s'aperçoit que, même si les techniques de prélèvement sont différentes, les résultats RT-PCR sont ressemblants (Fig. 48 et 49). L'expression du gène *PttGLIM1b* augmente graduellement des cellules méristématiques du cambium jusqu'au stade tardif de l'expansion cellulaire, puis diminue lors de la formation de la paroi secondaire jusqu'à la maturation tardive des cellules. Le même profil d'expression est mesuré pour le gène *PttWLIM1b* avec un niveau maximum de transcrite pendant la phase tardive de l'expansion des cellules du xylème. L'expression du gène *PttWLIM2b* augmente aussi avec

l'expansion cellulaire mais à un niveau plus faible et son expression chute très vite lors du processus de formation des parois secondaires (Hertzberg *et al.*, 2001). Ces données complétées des résultats RT-PCR indiquent que les gènes *PtGLIM1b*, *PtWLIM1b*, *PtPLIM2a*, *PtWLIM2a et b* et peut-être les gènes *PtGLIM1a* et *PtWLIM1a*, seraient impliqués dans l'élongation cellulaire des cellules du bois. Par ailleurs, ces gènes pourraient aussi jouer un rôle dans les étapes très précoces de la formation de la paroi secondaire qui nécessite une réorganisation massive du cytosquelette (Chaffey *et al.*, 2002) ainsi que dans la biosynthèse des précurseurs de la lignine tels que les phénylpropanoïdes (Kawaoka *et al.*, 2000). Une étude d'expression des protéines correspondantes et de leurs localisations dans les cellules du xylème éclaircirait cette question. De plus, la durée de vie des protéines LIM dans le bois est inconnue. En effet, ces protéines peuvent être très peu dégradées et ainsi s'accumuler dans les cellules du xylème formant la paroi secondaire après l'étape d'élongation cellulaire. D'autre part, par leurs fortes expressions du côté des tissus du phloème, les gènes *PtWLIM1a*, *PtβLIM1a* et *PtδLIM2b* pourraient être impliqués dans l'expansion des cellules du phloème ou l'élaboration de leur paroi cellulaire (Fig. 48). Mais actuellement, nous n'avons pas de données transcriptomiques détaillées sur l'expression des gènes en fonction des différentes étapes de formation du phloème, et la technique de prélèvement utilisée ici n'est pas assez précise pour différencier ces étapes.

Excepté le gène *PtGLIM1a* légèrement plus exprimé dans le xylème mature par rapport aux autres tissus, notamment le xylème en développement, l'expression des gènes *PtLIM* n'est globalement pas induite dans le xylème mature par rapport aux autres tissus de la tige (Fig. 48 et 49). Ces résultats concordent avec les données d'expression microarray comparant, chez *P. tremula × tremuloïdes*, l'expression des gènes entre le bois mort et le bois en début de formation (Moreau *et al.*, 2005). D'après cette étude, l'expression des gènes *PttWLIM2a et b* et *PttPLIM2a* diminue au moins d'un facteur trois dans les cellules de bois mort par rapport à celles de bois jeune en formation et l'expression du gène *PttWLIM1b* diminue d'un facteur six. L'expression du gène *PttGLIM1a* ne semble pas être modifiée dans le xylème entre les cellules jeunes et mortes, et celle du gène *PttGLIM1b* diminue légèrement n'excédant pas un facteur deux. Bien que ces gènes soient toujours exprimés dans le xylème mature, ces résultats indiquent que les gènes *PtLIM* ne sont peu ou pas impliqués dans les étapes les plus tardives de la formation du bois comme la mort cellulaire programmée. Cette interprétation est cohérente avec l'hypothèse que les protéines LIM de plante jouent un rôle dans la stabilisation des filaments d'actine et/ou la transcription des gènes (Kawaoka *et al.*, 2000 ; Thomas *et al.*, 2006). Les gènes *PtβLIM1a* et *PtδLIM2b* dont l'expression augmente

légèrement dans le xylème mature de bois normal par rapport au xylème en développement pourraient être impliqués dans la mort cellulaire programmée. Mais le fait qu'ils soient aussi exprimés avec la même intensité dans le cambium, le phloème et/ou l'écorce suggère qu'ils jouent un autre rôle dans le processus de différenciation cellulaire.

Par ailleurs, concernant le gène *PtGLIM1a*, des résultats d'analyses d'expression réalisées par cDNA-AFLP et Northern blot entre les zones de croissance primaire et secondaire de la tige (Prassinis *et al.*, 2005) sont identiques aux résultats RT-PCR obtenus dans cette étude (Fig. 48 et 49). Il a été montré chez le même clone INRA 717-1-B4 (*P. tremula* × *P. alba*), que le gène *PtGLIM1a* a un plus faible niveau d'expression dans la portion de tige proche de l'apex à croissance primaire que dans la zone de transition croissance primaire/secondaire, et la portion base de la tige à croissance secondaire où le phloème et le xylème sont lignifiés. De plus, le gène *PtGLIM1a* est fortement exprimé dans le xylème (en développement et mature) et est plus faiblement exprimé dans l'écorce (contenant du phloème et du cambium) et les feuilles matures (Prassinis *et al.*, 2005). Ce profil d'expression est en accord avec un possible rôle de ce gène dans la régulation des gènes de la voie des phénylpropanoïdes (Kawaoka *et al.*, 2000). De plus, de par sa forte expression dans le xylème mature, son rôle dans l'élaboration de la paroi secondaire est fort probable.

### 3.2.3. Les gènes *PtLIM* sont également exprimés dans les tissus vasculaires de la racine

Les gènes *PtLIM*, à l'exception du gène *PtβLIM1b*, sont aussi exprimés dans les jeunes racines et celles plus âgées formant des tissus vasculaires secondaires (Fig 48. et 49.). Deux profils d'expression différents se distinguent au niveau des tissus vasculaires de la racine de peuplier. Les gènes *PtWLIM1a*, *PtβLIM1a* et *PtWLIM2a* sont plutôt exprimés constitutivement dans tout les tissus de la racine (épiderme/phloème, zone cambiale, et xylème en développement et mature). Par contre, les gènes *PtGLIM1a* et *b*, *PtWLIM1b* et *PtPLIM2a* sont préférentiellement exprimés dans le xylème en développement de la racine. A ma connaissance, c'est la première fois qu'est réalisé une analyse d'expression de gènes au niveau des tissus vasculaires secondaires de la racine chez les arbres. Nous n'avons donc pas à notre disposition d'autres données d'expression de gènes *LIM* dans les tissus racinaires de peuplier. Ces deux types de profils (constitutif et tissus vasculaires du xylème) ont également été observés pour les mêmes gènes au niveau du tronc où des tiges de peuplier. Ainsi, les gènes *PtGLIM1a* et *b*, *PtWLIM1b* et *PtPLIM2a* sont spécifiques des tissus vasculaires formant



le bois à la fois au niveau des racines et au niveau des tiges, soulignant probablement leur importance dans la formation du xylème secondaire de la plante. On peut noter que le gène *PtWLIM1b* est aussi fortement exprimé dans le xylème en développement que dans la zone cambiale de la racine. Ce gène *PtWLIM1b* aurait peut-être une contribution plus importante dans les étapes très précoces de la xylogénèse que les gènes *PtGLIM1a* et *b*, et *PtPLIM2a*. Les gènes *PtWLIM1a*, *PtβLIM1a* et *PtWLIM2a* exprimés de manière ubiquitaire seraient plutôt impliqués dans des fonctions cellulaires essentielles pour la plante.

### 3.2.4. Les gènes *PtLIM* joueraient un rôle dans la formation du bois de tension

L'analyse de l'expression des gènes *PtLIM* par RT-PCR semi quantitative entre le bois tendu et le bois opposé, révèle que l'expression d'une grande proportion de gènes est plus ou moins induite dans les différents tissus de bois de tension (Fig. 50). Les gènes les plus fortement induits sont les gènes *PtGLIM1a* et *b* tous les deux aux niveau du xylème en développement et du xylème mature, le gène *PtWLIM1b* du phloème jusqu'au xylème en développement, et le gène *PtPLIM2a* au niveau du xylème en développement uniquement. Lorsque l'on regarde plus précisément les résultats récents d'une étude microarray analysant l'expression des gènes entre le bois normal et le bois tendu chez des peupliers (*P. tremula* L.) cultivés en plein champ (Andersson-Gunneras *et al.*, 2006), on s'aperçoit qu'ils sont en accord avec nos résultats RT-PCR. Dans le xylème en développement de bois de tension au stade de la formation de la couche G, les gènes *PtGLIM1a* et *b* sont les plus induits avec un ratio moyen de 1,8 entre le niveau d'expression dans le bois de tension et celui dans le bois normal. L'expression des gènes *PtPLIM2a* et *PtWLIM1b* est également induite d'un facteur 1,59 et 1,28 respectivement, tandis que l'expression des gènes *PtWLIM2a* et *b* diminue légèrement (ratio de 0,93 et 0,86 respectivement). Cette analyse d'expression démontre pour la première fois l'induction simultanée de plusieurs gènes *LIM* dans le bois de tension de peuplier. Bien que leurs profils d'expression dans les différents tissus de bois de tension ne soient pas exactement les mêmes, ce résultat suggère une redondance fonctionnelle des protéines à domaine LIM dans ce bois. Le bois de tension, de par la présence d'une couche gélatineuse supplémentaire, est un tissu enrichi en cellulose et contenant moins de lignines (Timell, 1969). Il est ainsi difficile de concevoir un rôle des gènes *PtGLIM1a* et *b*, *PtPLIM2a* et *PtWLIM1b* dans l'activation de la transcription des gènes *PAL*, *4CL* et *CAD* impliqués dans la biosynthèse des lignines et dans l'augmentation de la teneur en lignine, comme il a été

suggéré pour le gène *NtWLIM1* de tabac (Kawaoka *et al.*, 2000). De plus, il a été montré que l'expression des gènes impliqués dans la biosynthèse des lignines, incluant le gène *CAD* et certains gènes *4CL* et *PAL*, est le plus souvent réprimée dans le bois de tension (Pilate *et al.*, 2004a ; Andersson-Gunneras *et al.*, 2006). L'annotation de ces séquences n'étant pas suffisamment précise, ces données d'expression sont donc à prendre avec précaution. En effet, ces gènes appartiennent à des familles multigéniques, et certaines isoformes peuvent être impliquées dans d'autres processus que la lignification (Raes *et al.*, 2003 ; Tsai *et al.*, 2006 ; Tuskan *et al.*, 2006). A l'inverse, ces gènes *LIM* préférentiellement exprimés dans le bois de tension pourraient inhiber l'activation des gènes *PAL*, *4CL* et *CAD*. Ou alors, la fonction de ces protéines dans l'activation de la transcription serait inhibée par d'autres protéines lors de modifications post-traductionnelles tels que des phosphorylations ou déphosphorylations. J'ai en effet identifié pour les protéines PtGLIM1a et b, cinq sites putatifs de phosphorylation (Fig. 45). A l'avenir, il sera nécessaire de tester la capacité de fixation à l'ADN et d'identifier les séquences régulatrices cibles de ces facteurs de transcription LIM putatifs, afin de vérifier leur possible implication dans la régulation de la voie de biosynthèse des lignines. Enfin, les protéines LIM sous l'effet de stress mécanique pourraient être, suite à des modifications post traductionnelles ou des interactions protéine/protéine, spécifiquement dirigées vers le cytoplasme empêchant ainsi toute activité de transcription. Il a en effet été montré que la protéine CRP2 se déplace fortement vers les fibres de stress d'actine sous l'effet de l'étirement mécanique des cellules (Kim-Kaneyama *et al.*, 2005). De même, la localisation subcellulaire des protéines LIM de peuplier pourrait dépendre de l'état de différenciation des cellules du xylème. En effet, il a été montré que les protéines CRP2 et CRP3 ont une localisation plutôt nucléaire dans les cellules indifférenciées et cytoplasmique dans les cellules différenciées (Arber *et al.*, 1994 ; Arber & Caroni, 1996 ; Chang *et al.*, 2003). Ainsi, les protéines LIM de peuplier pourraient être localisées dans le noyau des initiales fusiformes en différenciation pour activer la transcription des gènes impliqués dans la lignification, et dans le cytoplasme des fibres ou éléments de vaisseaux matures pour le maintien des câbles d'actine intervenant dans la réponse aux contraintes mécaniques. Par ailleurs, il a été montré que la taille des fibres et des vaisseaux du xylème pourrait être sensiblement plus longue dans le bois de tension (Jourez *et al.*, 2001). Etant donné que les protéines PLIM1 ou PLIM2 pourraient intervenir dans la croissance du tube pollinique (Baltz *et al.*, 1992b ; Baltz *et al.*, 1999 ; Sweetman *et al.*, 2000), l'hypothèse d'un rôle des protéines PtGLIM1a et b, PtPLIM2a et PtWLIM1b dans la croissance accrue des fibres ou des vaisseaux de bois de tension n'est pas exclue. D'autre part, la couche G est quasiment constituée exclusivement de cellulose

dont l'angle des microfibrilles, différent de celui des autres couches S1 et S2 de la paroi secondaire, est presque parallèle à l'axe de la fibre. De la même manière, les microtubules arrangés hélicoïdalement dans les fibres de bois normal, sont arrangés axialement dans les fibres de bois de tension. Il a ainsi été suggéré un éventuel rôle du réseau de microtubules dans l'organisation de l'orientation des microfibrilles de cellulose dans le bois normal et tendu (Chaffey *et al.*, 2002 ; Pilate *et al.*, 2004b). De plus, une étude récente montre une association fonctionnelle entre les microtubules et la cellulose synthase (CESA) chez *Arabidopsis* (Paredez *et al.*, 2006). Enfin, une colocalisation entre la protéine HaWLIM1 et les microtubules a été montrée dans les protoplastes de tournesol (Brière *et al.*, 2003), et de manière intéressante, les cellulose synthases possèdent à leur face cytoplasmique un domaine riche en cystéine similaire aux domaines RING à doigt de zinc et permettant la dimérisation de ces protéines (Kurek *et al.*, 2002). Ainsi, les protéines PtGLIM1a et b, PtPLIM2a et/ou PtWLIM1b pourraient agir comme adaptateur moléculaire entre les microtubules et certaines cellulose synthases induites lors de la formation du bois de tension (Wu *et al.*, 2000 ; Joshi, 2003). Or, l'interaction de la protéine HaWLIM1 avec les microtubules est sujette à controverse du fait de la découverte de l'interaction directe de la protéine NtWLIM1 avec les filaments d'actine (Thomas *et al.*, 2006). De plus, la protéine NtWLIM1 empaquette les filaments d'actine en des câbles épais. Les filaments d'actine ont une orientation axiale dans les fibres de bois normal et celles de bois de tension, et aucune différence dans l'épaisseur des filaments d'actine n'a été signalée (Chaffey *et al.*, 2002). Néanmoins, un rôle des protéines LIM de peuplier dans la régulation de l'épaisseur des filaments d'actine dans les fibres G de bois de tension est attrayant. En effet, ces câbles d'actine offrent peut-être une résistance à la contraction ou l'étirement des fibres, et la stabilisation des câbles d'actine peut aussi faciliter le flux cytoplasmique permettant le transport vésiculaire de matériel nécessaire à l'élaboration de l'épaisse couche gélatineuse. Par ailleurs, les microtubules se trouvent co-alignés avec les microfilaments d'actine dans les fibres de bois tendu (Chaffey *et al.*, 2002). Ces deux éléments du cytosquelette sont approximativement alignés parallèlement aux striations des microfibrilles de cellulose dans la paroi des fibres G. Ces quatre protéines LIM pourraient jouer un rôle dans l'organisation du cytosquelette des fibres de bois de tension en participant à un complexe moléculaire liant à la fois les microtubules, les cellulose synthases et les filaments d'actines. Enfin, l'hypothèse d'un rôle des protéines PtLIM dans la perception ou la transduction du signal généré par un stress mécanique est à envisager. En effet, chez les animaux, la protéine CRP1 co-localise avec l' $\alpha$ -actinine le long des fibres de stress d'actine des cellules musculaires lisses et des fibroblastes d'embryons de poulet ou de rat (Pomies *et*

*al.*, 1997 ; Tran *et al.*, 2005). Ces fibres de stress d'actine sont des câbles de filaments d'actine contractiles attachés aux plaques d'adhésion focale et intervenant dans la contraction des cellules animales. De plus, des études préliminaires ont montré que les protéines fusions NtWLIM1-GFP et NtWLIM2-GFP de tabac exprimées dans les fibroblastes humains ou les cellules simiennes de rein, colocalisent avec fibres de stress d'actine et s'accumulent au niveau des plaques d'adhésion focale (Steinmetz *et al.*, 2006). Les protéines LIM chez le peuplier pourraient ainsi jouer un rôle dans les mécanismes encore inconnus de tension/contraction cellulaire dans le bois. Cette hypothèse est d'autant plus vraisemblable qu'il a été démontré une implication des protéines CRP2 et CRP3 animales dans des mécanismes de réponse au stress mécanique comme l'étirement des cellules (Knoll *et al.*, 2002 ; Kim-Kaneyama *et al.*, 2005). Une localisation sub-cellulaire des protéines PtGLIM1a et b, PtPLIM2a et PtWLIM1b ainsi que l'identification de leurs partenaires protéiques et nucléiques, apporteraient des éléments de réponse à toutes ces questions.

### 3.2.5. Les gènes *PtPLIM2a* et *b* et *PtδLIM2a* et *b* sont spécifiquement exprimés dans le pollen

Concernant la spécificité de l'expression de certains gènes *PtLIM* dans le pollen, les résultats obtenus concordent avec la classification phylogénétique établie précédemment. En effet, les gènes *PtPLIM2a* et *b*, classés dans le sous-groupe PLIM2 spécifique du pollen, sont très fortement exprimés dans le pollen, comme les gènes *AtPLIM2a*, *NtPLIM2* et *HaPLIM2* (Eliasson *et al.*, 2000 ; Becker *et al.*, 2003 ; Honys & Twell, 2003). La forte expression des gènes *PtδLIM2a* et *b* dans le pollen et les étamines est un résultat intéressant. En effet, les deux gènes *PtδLIM2a* et *b* et les gènes *δLIM2* des espèces de la sous classe des astérides (solanacées et astéracées) sont encadrés par les sous groupes PLIM2 des dicotylédones et monocotylédones, suggérant également pour ces gènes une forte expression dans le pollen. Mais les origines des ESTs correspondant aux gènes *δLIM2* chez les astérides ne nous donnaient guère d'informations sur leur expression dans le pollen. Avec un profil d'expression similaire aux gènes *PtPLIM2a* et *b*, nous avons ici la confirmation que les gènes *PtδLIM2a* et *b* sont aussi fortement exprimés dans le pollen et peuvent avoir une fonction redondante avec celle des gènes *PtPLIM2a* et *b*.

L'expression des gènes *PtPLIM2a* et *b* et *PtδLIM2a* et *b* augmente en fonction du développement des étamines (Fig. 51). On ne peut exclure la possibilité que ces transcrits détectés dans les étamines soient produits par les parois ou les filaments des anthères ou

encore les tissus vasculaires des étamines. Cependant, il est fort probable, en regard de la forte expression des gènes *PtPLIM2a* et *b* et *Pt $\delta$ LIM2a* dans le pollen isolé, que les signaux observés dans les étamines soient plutôt dus à une expression de ces gènes dans les gamètes mâles en cours de développement. Une expression dans les anthères est également observée pour les gènes *PLIM1* et *PLIM2* chez *Gerbera hybrida* et le tabac (Eliasson *et al.*, 2000 ; Sweetman *et al.*, 2000 ; Laitinen *et al.*, 2005). Chez *Arabidopsis*, les transcrits *AtPLIM2a*, *b* et *c* s'accumulent préférentiellement dans les anthères à un stade tardif de développement (Amagai *et al.*, 2003 ; Honys & Twell, 2004). Le niveau d'expression de ces gènes augmente au moins d'un facteur 10 entre le stade de microspore mononuclée et le stade de grain de pollen tricellulaire et mature. Dans les anthères de tournesol, la protéine HaPLIM1 est détectée par immunohistochimie dans les grains de pollen immature au stade microspore mononuclée, dans le pollen mature, mais pas dans les grains de pollen au stade bicellulaire (Baltz *et al.*, 1999). Comme tous les gènes du groupe  $\delta$ LIM2 chez le peuplier sont aussi fortement exprimés dans le pollen mature, il est probable qu'au cours du développement du pollen dans les étamines, les grains de pollen accumulent ces protéines pour anticiper la future formation du tube pollinique.

L'accumulation de transcrits *PtPLIM2a* et *b* et *Pt $\delta$ LIM2a* et *b* est aussi très forte dans des grains de pollen germés présentant des tubes polliniques, suggérant leurs implications dans la croissance du tube pollinique chez le peuplier (Fig. 51). Avec la forte accumulation de la protéine HaPLIM1a dans les cônes germinatifs des grains de pollen trinuéclées, il a également été suggéré que la protéine HaPLIM1a participe à la germination du pollen ou la croissance du tube pollinique (Baltz *et al.*, 1999). La protéine HaPLIM1 co-localise avec l'actine dans les cônes germinatifs, mais une interaction directe entre les deux protéines n'a pas encore été mise en évidence. Par ailleurs, l'actine forme un réseau de filaments, alors que HaPLIM1 s'accumule en de petites taches corticales ressemblant à des sites d'attachement à la membrane plasmique. Il a ainsi été proposé une possible interaction entre la protéine HaPLIM1, le cytosquelette d'actine et la membrane plasmique (Baltz *et al.*, 1999). Toutefois, la localisation de la protéine HaPLIM1a dans les tubes polliniques n'a pas été testée, car il est actuellement très difficile de faire germer du pollen de tournesol *in vitro*. Les gènes *NtPLIM1a* et *b* chez le tabac sont aussi exprimés dans les pollens germés, mais moins fortement que dans le pollen mature (Sweetman *et al.*, 2000). Sachant que le cytosquelette d'actine est indispensable à la formation du tube pollinique des plantes (Taylor & Hepler, 1997 ; Chen *et al.*, 2006) et qu'il a été récemment mis en évidence que la protéine NtWLIM1 interagit directement avec l'actine filamenteuse, stabilise et empaquette les filaments d'actine

(Thomas *et al.*, 2006), les protéines PtPLIM2a et b et Pt $\delta$ LIM2a et b pourraient être impliquées dans la croissance du tube pollinique en stabilisant la formation de câbles d'actine épais.

Les gènes *PLIM2* et  *$\delta$ LIM2* chez le peuplier ne sont pas exclusivement exprimés dans le pollen ou les étamines. Cette expression non pas spécifique, mais plutôt préférentielle des gènes *PLIM2* dans le pollen a déjà été observée pour d'autres gènes tels que *AtPLIM2*, *ZmPLIM2a* et *OsPLIM2c* (Eliasson *et al.*, 2000 ; Yoshida *et al.*, 2005). Le cas le plus intéressant chez le peuplier est l'induction de l'expression du gène *PtPLIM2a* dans le xylème en développement de bois de tension (Fig. 50). Cette accumulation de transcrits à la fois dans le pollen germé et dans le xylème tendu montre que le gène *PtPLIM2a* est probablement impliqué dans un ou des processus cellulaires communs intervenant dans ces deux tissus très contrastés. Ce peut-être à la fois l'élongation cellulaire, la perception ou la réponse à la contrainte mécanique, ou encore la biosynthèse de cellulose. Chez de jeunes peupliers, le gène *Pt $\delta$ LIM2a* est exclusivement exprimé dans le phloème et l'écorce et son expression augmente avec le stress mécanique. De même, le gène *Pt $\delta$ LIM2b* semble répondre à la contrainte mécanique. Son expression augmente dans le phloème/écorce de la face supérieure de la tige inclinée (Fig. 50). Les tubes polliniques subissent également un stress mécanique dû à la pression de turgescence lors de la croissance du tube, ou à la résistance mécanique lors de la pénétration du stigmate et du style (Steer & Steer, 1989 ; Geitmann & Parre, 2004). Une faible expression des gènes *PtPLIM2b*, et *Pt $\delta$ LIM2a* et *b* est aussi observée au niveau des stigmates des fleurs femelles. Les rameaux prélevés sur des peupliers naturels, comportaient certains bourgeons avec des fleurs femelles déjà ouvertes à leurs extrémités (Dowkiw A. Com. personnelle). Nous ne pouvons alors exclure une possible pollinisation de ces fleurs avant le prélèvement, et le signal observé proviendrait de grain de pollen germés avec des tubes polliniques traversant les stigmates.

### 3.2.6. Les gènes *PtLIM* interviennent dans la formation des fibres cotonneuses chez le peuplier

Pour les gènes *PtWLIM1a*, *Pt $\beta$ LIM1a* et *b*, *PtPLIM2a* et *b*, et *Pt $\delta$ LIM2a* et *b*, on observe une expression dans les fibres cotonneuses et les capsules de peuplier. Chez le cotonnier, les fibres de coton sont des trichomes unicellulaires dérivant des cellules épidermiques de l'ovaire et entourant la graine. Elles subissent au cours de leurs développement une phase d'élongation cellulaire considérable, suivie d'une phase de synthèse

de la paroi secondaire. La paroi des fibres de coton est composée à 95 % de cellulose et semble dépourvue de lignine (Ruan, 2007). Le développement des fibres cotonneuses chez le peuplier n'a jamais été étudié. D'après nos observations et contrairement aux fibres de coton, les fibres chez le peuplier ne sont pas attachées aux graines mais proviennent de la base de la capsule. Chez le peuplier, les fibres cotonneuses sont issues après fécondation des cellules du placenta supportant le funicule et l'ovaire (Boes & Strauss, 1994). Le type cellulaire de ces fibres est inconnu, mais il est fort probable que ces cellules épidermiques se soient aussi différenciées en trichomes unicellulaires de la même manière que les fibres de cotonnier. On constate une forte expression de ces gènes *PtLIM* dans les capsules. Au vu de l'aspect très chlorophyllien de la capsule et de sa structure visant à protéger les graines lors de leurs maturations, il est difficile de concevoir des processus d'élongation cellulaire similaires à celui fibres cotonneuses dans les cellules de la capsule. Or, quand les capsules sont mures et relâchent leurs graines, chaque graine est accompagnée par une portion de placenta portant plusieurs fibres cotonneuses (Boes & Strauss, 1994). Lors de la récolte des graines et fibres, la majorité des capsules n'étaient pas mures et ouvertes. Il est alors possible que les corps cellulaires des fibres cotonneuses unicellulaires soient restés fixés aux capsules lors du prélèvement des fibres. De plus, chez le cotonnier, seulement 25-30 % des cellules épidermiques se différencient en fibre et la majorité des cellules se développent en des fibres courtes ou duvet (Ruan, 2007). Peut-être qu'au moment de la récolte, la majorité des fibres synthétisées par la capsule étaient trop courtes pour être visible et prélevée en tant que fibre. Ces difficultés techniques pourraient expliquer le fait que l'on retrouve une plus grande quantité de transcrits *PtLIM* dans les capsules plutôt que dans les fibres. Toutefois ces résultats suggèrent un rôle des gènes *PtWLIM1a*, *PtβLIM1a*, *PtPLIM2a* et *b* et *PtδLIM2a* et *b* dans le développement et la croissance des fibres cotonneuses de peuplier. La suppression par RNAi du gène *ACTINI* chez le coton, montre que l'actine est indispensable à l'élongation des fibres de coton (Li *et al.*, 2005). La forte expression de certains gènes *LIM* dans les fibres cotonneuses est une nouvelle fois un exemple d'une possible implication des gènes *LIM* dans l'élongation cellulaire, via le cytosquelette d'actine. Par ailleurs, ces fibres cotonneuses sont enrichies en cellulose tout comme le bois de tension, montrant également un éventuel rôle des protéines LIM dans le contrôle de la biosynthèse de la cellulose. Toutefois, la forte expression des gènes *PtPLIM2a* et *b* et *PtδLIM2a* et *b* dans les fibres et capsules peut paraître surprenante dans l'hypothèse où les gènes du groupe  $\delta$ LIM2 interviendraient dans des processus d'élongation cellulaire polaire et localisée du tube pollinique. En effet, l'élongation cellulaire de la fibre de coton n'est pas anisotrope et localisée à l'extrémité de la fibre,

mais est plutôt diffuse le long de la fibre (Ruan, 2007). Par ailleurs, il a été suggéré une participation des protéines PLIM1 et PLIM2 dans le maintien des longs câbles d'actine présents dans le tube pollinique (Thomas *et al.*, 2006). Les fibres de coton en développement possèdent également d'épais filaments d'actines ayant une orientation longitudinale, surtout à trois jours après l'anthèse (Li *et al.*, 2005). Ainsi, les protéines à domaine LIM pourraient aussi avoir un rôle dans l'organisation particulière du cytosquelette d'actine de ces cellules très allongées. De façon inattendue, les gènes *PtGLIM1a* et *b* pourtant phylogénétiquement les plus proches du gène *GhaLIM1a* de cotonnier contenant le plus d'ESTs dans les fibres de coton, ne sont pas du tout exprimés dans les fibres ou des capsules de peuplier. Dans le cas où *PtGLIM1a* et *b* répondraient effectivement à une contrainte mécanique, l'absence d'expression de ces gènes dans les fibres cotonneuses n'est pas étonnante car ce tissu protégé par la capsule subit a priori moins de stress mécaniques comparé au bois de tension.

### 3.2.7. Les gènes dupliqués chez le peuplier

Les analyses de séquence des douze gènes de peuplier codant pour des protéines à deux domaines LIM nous montrent qu'ils forment tous des paires de gènes suggérant un événement de duplication récent. La signification fonctionnelle et évolutive de ces événements récents de duplication mérite d'être élucidée. En effet, la duplication d'un gène qui est un événement aléatoire n'apporte pas forcément d'avantages sélectifs à la plante, et le gène dupliqué peut être contre sélectionné au cours de l'évolution (Lynch & Conery, 2000). Toutefois, les résultats RT-PCR montrent que les gènes *PtLIM* sont tous exprimés chez le peuplier, indiquant que les gènes dupliqués sont sûrement fonctionnels. Des paires de gènes codant pour des cellulose synthases (*CesA*) ont également été découvertes chez *P. trichocarpa* et pour au moins une paire, les deux copies sont activement transcrites (Djerbi *et al.*, 2005). La conservation des paires de gènes *CesA* dupliqués n'est pas étonnante du fait de l'importance des cellulose synthases dans l'élaboration de la paroi des fibres de bois. Les gènes *PtLIM* ont tous été conservés au cours de l'évolution parce qu'ils doivent être également importants dans le développement du peuplier. Les gènes *PtLIM* dupliqués ayant une similitude de séquence très forte (85 à 95 % d'identité entre eux) ont a priori des fonctions très voisines ou similaires. L'expression des gènes *PtLIM* n'est pas régulée de la même manière entre les gènes dupliqués. La duplication des gènes *PtLIM* a peut-être permis à l'une ou l'autre copie d'une paire d'acquérir une nouvelle fonction supplémentaire (Lynch &



Conery, 2000). Dans le cas des gènes *PtWLIM1a* et *b*, le gène *PtWLIM1a* exprimé constitutivement dans l'arbre pourrait intervenir dans des fonctions cellulaires essentielles au développement de la plante, et ne semble pas être impliqué dans des processus spécifiques à la physiologie de l'arbre. L'importance des protéines WLIM1 peut-être illustrée par l'impossibilité de produire des lignées transgéniques de tabac surexprimant constitutivement la protéine NtWLIM1-GFP sous le contrôle du promoteur 35S (Hoffmann *et al.*, 2005). De plus, chez *Arabidopsis*, l'homozygotie pour l'allèle d'insertion ADN-T du gène *AtWLIM2a* semble être létale pour la plante au stade embryonnaire (Dean Rider *et al.*, 2003). Ainsi, on peut également envisager que la perte de l'expression du gène *PtWLIM1a* dans certains tissus n'est pas viable. Seul une copie de ce gène pourrait avoir une expression spécifique dans un tissu particulier. Il semblerait que ce soit le cas du gène *PtWLIM1b* exclusivement exprimé dans les tissus vasculaires de peuplier. Ce gène aurait donc vraisemblablement acquis une fonction dans le développement vasculaire de l'arbre. De plus, les gènes dupliqués *PtGLIM1a* et *b* sont aussi spécifiquement exprimés dans les tissus vasculaires. Ces gènes seraient probablement apparus après une duplication ancienne d'un gène ancestral *WLIM1* et se seraient spécialisés dans certains processus spécifiques du xylème. Nous pouvons aussi noter que, par rapport au gène *PtGLIM1b*, le gène *PtGLIM1a* est plus spécifiquement exprimé lors des étapes tardives de la différenciation du xylème secondaire. Un autre exemple marquant est l'expression quasiment constitutive de *PtβLIM1a* dans la plupart des tissus analysés, alors que son dupliquât *PtβLIM1b* n'est exprimé que dans les organes reproducteurs. La duplication du gène *PtβLIM1a*, a pu permettre l'apparition d'une fonction spécifique dans les inflorescences de peuplier. Cette fonction encore inconnue serait portée par le gène *PtβLIM1b*. Par contre, les gènes *PtPLIM2a* et *b* et *PtδLIM2a* et *b* semblent eux avoir des fonctions redondantes et essentielles au niveau du pollen. Toutefois, dans d'autres tissus, l'expression de ces gènes n'est pas chevauchante indiquant peut-être des spécificités de fonction particulières.

## 4 Conclusions

Cette analyse de l'expression des gènes *LIM* de peuplier par RT-PCR semi quantitative a confirmé les hypothèses issues de l'étude phylogénétique sur les profils d'expression des gènes *PLIM2*, *WLIM1* et *WLIM2*. En effet, les gènes *PtPLIM2a* et *b* sont préférentiellement exprimés dans le pollen et les gènes *PtWLIM1a* et *PtWLIM2a* ont plutôt une expression constitutive dans la plante. Par ailleurs, les profils d'expression des nouveaux gènes *LIM* identifiés ont été déterminés. Ainsi, au sein du groupe  $\beta$ LIM1, le gène *Pt $\beta$ LIM1a* semble exprimé constitutivement et l'expression du gène *Pt $\beta$ LIM1b* est spécifique des organes reproducteurs excepté le pollen mature. De plus, les gènes  $\delta$ LIM2, phylogénétiquement proches des gènes *PLIM2* sont aussi préférentiellement exprimés dans les étamines et le pollen. Cette étude a mis en évidence de nouvelles spécificités d'expression de certains gènes *LIM*, notamment les gènes *PtGLIM1a* et *b*, dans les tissus vasculaires primaires et secondaires de peuplier. De même, la duplication récente des gènes *LIM* de peuplier a permis au gène *PtWLIM1b* d'acquérir une expression spécifique des tissus vasculaires. Cette analyse d'expression de gène nous montre aussi qu'en réalité un nombre conséquent de gènes *LIM* ont une expression plus ou moins induite dans le bois de tension par rapport au bois opposé. Ce profil d'expression particulier soulève des questions sur le rôle des protéines LIM dans le contrôle de la lignification des cellules. Enfin, la forte expression de certains gènes *LIM* dans les fibres cotonneuses de peuplier suggère une fonction générale de ces gènes dans certains aspects de l'élongation cellulaire.

Afin de poursuivre la caractérisation fonctionnelle d'un gène *LIM* de peuplier, nous avons choisi de concentrer nos efforts sur le gène *PtGLIM1a*. En effet, c'est le gène dont l'expression semble la plus fortement induite dans le bois de tension, et plus particulièrement à un stade tardif de différenciation du xylème pouvant correspondre à l'étape de synthèse de paroi secondaire enrichie en cellulose. La thématique de recherche de l'équipe est notamment axée sur la compréhension des mécanismes régulant la synthèse de la paroi secondaire des fibres de xylème de bois tendu. Pour continuer l'étude fonctionnelle de la protéine *PtGLIM1a*, nous avons décidé dans un premier temps de la produire dans la bactérie sous forme de protéine recombinante afin d'obtenir une protéine purifiée et des anticorps dirigés contre elle.

## Partie 3. Approche biochimique de l'étude de la protéine PtaGLIM1a

### 1 Matériels et méthodes

#### 1.1. Souches de bactéries utilisées

Pour le clonage de la séquence codante *PtaGLIM1a* dans les vecteurs du système Gateway<sup>®</sup>, deux souches de bactéries *Escherichia coli* ont été utilisées **DB3.1** (Invitrogen<sup>™</sup>) de génotype [ $F^-$  *gyrA462 endA1*  $\Delta$ (*sr1-recA*) *mcrB mrr hsdS20*( $r_B^-$ ,  $m_B^-$ ) *supE44 ara14 galK2 lacY1 proA2 rpsL20*(S<sub>mr</sub>) *xyl5*  $\Delta$ *leu mtl1*] et résistante au produit du gène *ccdB* et **JM109** (Promega<sup>®</sup>) de génotype [ $e14^-$ ( $McrA^-$ ) *recA1 endA1 gyrA96 thi-1 hsdR17* ( $r_K^- m_K^+$ ) *supE44 relA1*  $\Delta$ (*lac-proAB*) [ $F'$  *traD36 proAB lacI<sup>q</sup>Z* $\Delta$ *M15*]] se transformant facilement et sensible au produit du gène *ccdB*. Cette souche a également été utilisée pour le clonage des produits PCR dans le vecteur pGEM<sup>®</sup>-T Easy.

Pour la production de protéines recombinantes, différentes souches de *E. coli* ont été utilisées : La souche **BL21(DE3)** (Novagen<sup>®</sup>) de génotype [ $F^-$  *ompT hsdS<sub>B</sub>* ( $r_B^- m_B^-$ ) *gal dcm* (DE3)] est la plus couramment utilisée avec la souche **BL21(DE3) pLysS** (Novagen<sup>®</sup>) de génotype [ $F^-$  *ompT hsdS<sub>B</sub>* ( $r_B^- m_B^-$ ) *gal dcm* (DE3) pLysS (Cam<sup>R</sup>)]. Le plasmide pLysS produit le lysozyme T7 qui supprime l'expression basale de l'ARN polymérase T7 avant induction, ce qui est utile dans le cas d'une protéine toxique pour la bactérie. Il permet également une meilleure lyse des cellules lors de l'extraction des protéines.

Les souches **Rosetta<sup>™</sup>(DE3)** (Novagen<sup>®</sup>) de génotype [ $F^-$  *ompT hsdS<sub>B</sub>*( $r_B^- m_B^-$ ) *gal dcm* (DE3) pRARE (Cam<sup>R</sup>)], et **Rosetta<sup>™</sup> 2(DE3) pLysS** (Novagen<sup>®</sup>) de génotype [ $F^-$  *ompT hsdS<sub>B</sub>*( $r_B^- m_B^-$ ) *gal dcm* (DE3) pLysSRARE2 (Cam<sup>R</sup>)] sont utilisées pour l'expression de protéines eucaryotes contenant beaucoup de codons rares pour *E. coli*. Le plasmide pRARE contient les gènes d'ARNt *argU, argW, ileX, glyT, leuW, proL, metT, thrT, tyrU, et thrU*, et les codons rares AGG, AGA, AUA, CUA, CCC, et GGA sont ainsi complétés. Le plasmide pRARE2 contient en plus du plasmide pRARE, le gène d'ARNt *argX* qui reconnaît le codon rare CGG codant pour l'arginine.

Les souches **Origami(DE3)** (Novagen<sup>®</sup>) de génotype [ $\Delta$ (*ara-leu*)7697  $\Delta$ *lacX74*  $\Delta$ *phoA PvuII phoR araD139 ahpC galE galK rpsLF'*[*lac<sup>+</sup> lacI<sup>q</sup> pro*] (DE3) *gor522::Tn10 trxB*

### Partie 3. La protéine PtaGLIM1a

(Kan<sup>R</sup>, Str<sup>R</sup>, Tet<sup>R</sup>), et **Origami(DE3) pLysS** (Novagen<sup>®</sup>) de génotype [ $\Delta(ara-leu)7697 \Delta lacX74 \Delta phoA PvuII phoR araD139 ahpC galE galK rpsLF[lac^+ lacI^q pro]$  (DE3) *gor522::Tn10 trxB* pLysS (Cam<sup>R</sup>, Kan<sup>R</sup>, Str<sup>R</sup>, Tet<sup>R</sup>)], sont recommandées pour les protéines nécessitant des ponts disulfures pour leur repliement, car elles contiennent des mutations pour à la fois la thiorédoxine réductase (*trxB*) et la glutathione réductase (*gor*) favorisant ainsi la formation de ponts disulfures dans le cytoplasme.

La souche **Tuner(DE3) pLysS** (Novagen<sup>®</sup>) de génotype [ $F^- ompT hsdSB(r_B^- m_B^-) gal dcm lacY1$  (DE3) pLysS (CamR)] contient une mutation dans le gène lac perméase (*lacY*) permettant une entrée uniforme de l'IPTG dans toutes les cellules d'une population bactérienne et ainsi le niveau d'expression de la protéine peut être ajusté pour toutes les cellules de la culture.

La souche **C41** (OverExpress<sup>™</sup>) de génotype [ $F^- ompT gal hsdS_B (r_B^- m_B^-) dcm lon$  (DE3)] est un mutant non caractérisé de la souche BL21(DE3), et la souche **C43** (OverExpress<sup>™</sup>) de génotype [ $F^- ompT gal hsdS_B (r_B^- m_B^-) dcm lon$  (DE3)] est un double mutant non caractérisé dérivant de la souche C41 (Miroux & Walker, 1996). Ces deux souches produisent des membranes intracellulaires proliférantes, et permettent ainsi la surproduction de protéines membranaires ou globulaires (Miroux & Walker, 1996 ; Arechaga *et al.*, 2000).

#### 1.2. Préparation de bactéries compétentes pour la transformation par choc thermique

Une culture de cellules bactériennes de 10 ml est lancée en milieu LB liquide (*Luria-Bertani*) [Bacto<sup>®</sup>-tryptone 10 g/l ; extrait de bactéries et de levure 5 g/l ; NaCl 175 mM] à 37°C de façon à ce que la DO<sub>600nm</sub> initiale soit inférieure ou égale à 0,2 unité/ml. Quand la DO<sub>600nm</sub> est égale à 0,7-0,8 unité/ml, les cellules sont centrifugées à 1500 g pendant 10 min à 4°C. Le culot de bactéries est repris doucement dans 500 µl de milieu TSB froid [LB liquide, PEG (PM 3350 Da) 10 % ; DMSO 5 % ; MgCl<sub>2</sub> 10 mM ; MgSO<sub>4</sub> 10 mM]. Les cellules sont incubées au minimum 30 min et jusqu'à 24h à 4°C.

#### 1.3. Transformation par choc thermique des bactéries compétentes

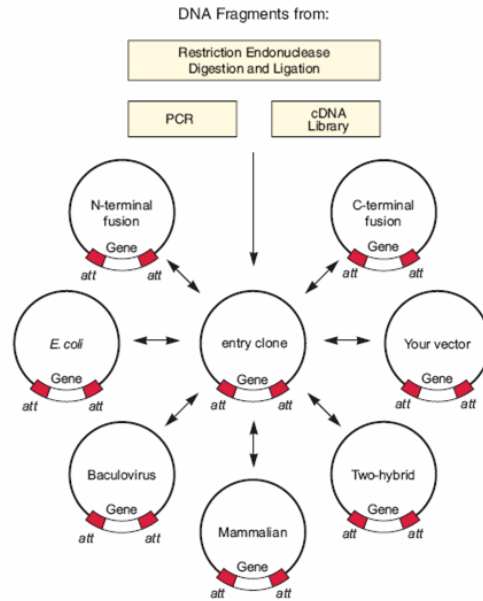
La transformation est réalisée sur 50 µl de bactéries compétentes dans un microtube préalablement refroidi dans la glace contenant 0,5 à 3 µl de plasmide. Le mélange est incubé

pendant 30 min dans la glace, puis un choc thermique de 30 secondes à 42°C est effectué pour la pénétration du plasmide dans les bactéries. Pour la régénération des bactéries transformées avant étalement, 950 µl de milieu LB est ajouté et les bactéries sont cultivées à 37°C pendant 1 heure sous agitation. Les bactéries sont ensuite étalées sur milieu LB solide (ajout d'agarose à 15 g/l) contenant des antibiotiques appropriés pour la sélection des transformants.

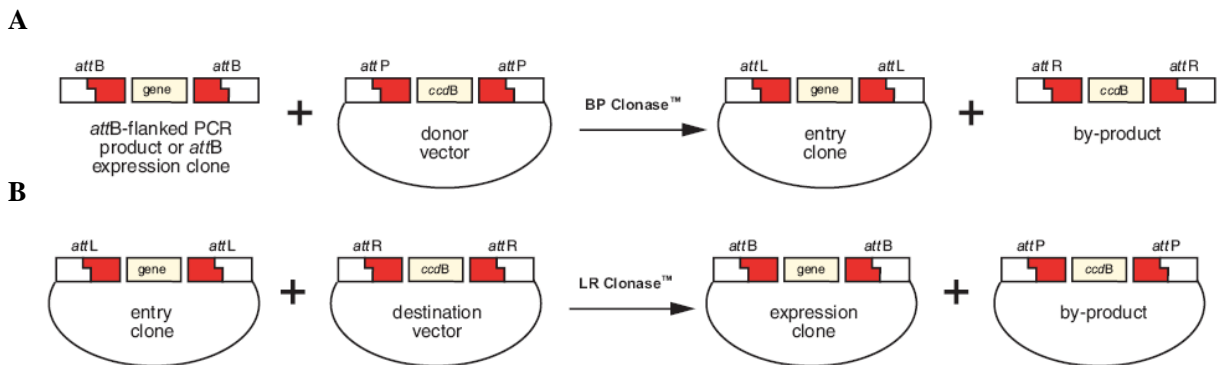
#### 1.4. Principe du clonage de gènes en utilisant le système Gateway®

Afin de cloner la séquence codante du gène *PtaGLIM1a* dans des vecteurs d'expression pour la production chez *E. coli* de protéines recombinantes PtaGLIM1a fusionnées à la Glutathione S-Transférase (GST) et une étiquette de six histidines (6His), nous avons utilisé le système Gateway®. Ce système de clonage est basé sur la création d'un vecteur d'entrée contenant notre gène d'intérêt. Ce vecteur d'entrée permettra par la suite d'insérer par recombinaison notre gène dans différents vecteurs de destination (Fig. 53). Cette réaction de recombinaison s'effectue grâce à des sites de recombinaison *attB*, *attP*, *attL* et *attR* et deux enzymes : la BP Clonase™ et la LR Clonase™ (Fig. 54A et B). De nombreux vecteurs de destination ont été créés pour des applications diverses telles que l'expression de protéines recombinantes dans *E. coli* et la levure, ou bien la production de plantes transgéniques surexprimant ou réprimant l'expression du gène d'intérêt. Pour la production de protéines recombinantes GST et 6His-PtaGLIM1a, nous avons choisi de cloner directement le produit PCR de notre gène d'intérêt dans un vecteur d'entrée pENTR-TOPO® linéarisé grâce à la topoisomérase I. Avec cette méthode, on obtient un vecteur d'entrée contenant notre gène flanqué de part et d'autre de sites de recombinaison *attL*. Avec le vecteur d'entrée ainsi créé, nous pouvons intégrer par recombinaison avec la LR Clonase™ notre gène d'intérêt dans différents vecteurs de destination portant les bordures *attR*. Nous pouvons ainsi générer différents clones d'expression. Le vecteur pENTR-TOPO® et les vecteurs de destination contiennent le gène *ccdB* entre les sites de recombinaison. Le gène *ccdB* est toxique pour les bactéries couramment utilisées en laboratoire, et seules les bactéries DB3.1, résistantes au produit de ce gène, sont utilisées pour la production des vecteurs d'entrée et de destination. Ce gène *ccdB* permet la sélection des bactéries JM109 qui ont intégré les vecteurs ayant effectivement remplacé par recombinaison le gène *ccdB* par notre gène d'intérêt.

### Partie 3. La protéine PtaGLIM1a



**Figure 53.** Illustration de l'avantage du système Gateway® par rapport à la méthode de clonage classique. La technologie Gateway® est une méthode basée sur la recombinaison site spécifique du bactériophage lambda (Landy, 1989). Cette technologie permet, d'une manière rapide et efficace, de cloner une séquence d'ADN dans de multiples vecteurs différents pour l'analyse fonctionnelle et l'expression de protéines (Hartley *et al.*, 2000).



**Figure 54.** Représentation schématique du principe du clonage et de la recombinaison dans les vecteurs du système Gateway®

(A) La réaction BP permet le clonage par recombinaison d'un produit PCR bordé par les sites de recombinaison *attB* (ou d'un clone d'expression *attB* linéarisé) avec un vecteur donneur possédant les sites de recombinaison *attP* pour créer un vecteur d'entrée contenant les sites de recombinaison *attL*. Cette réaction est catalysée par la BP clonase™.

(B) La réaction LR permet la recombinaison entre un clone d'entrée possédant le site de recombinaison *attL* et un vecteur de destination possédant les sites de recombinaison *attR* pour créer un clone d'expression contenant les sites de recombinaison *attB*. Cette réaction est catalysée par la LR Clonase™.

#### 1.5. Clonage de l'ADNc *PtaGLIM1a* dans le vecteur *pENTR™/D-TOPO®* avec le système Gateway®

La séquence codante de l'ADNc *PtaGLIM1a* a été amplifiée par PCR à partir du vecteur purifié correspondant au clone EST CF235176, et en utilisant les amorces sens

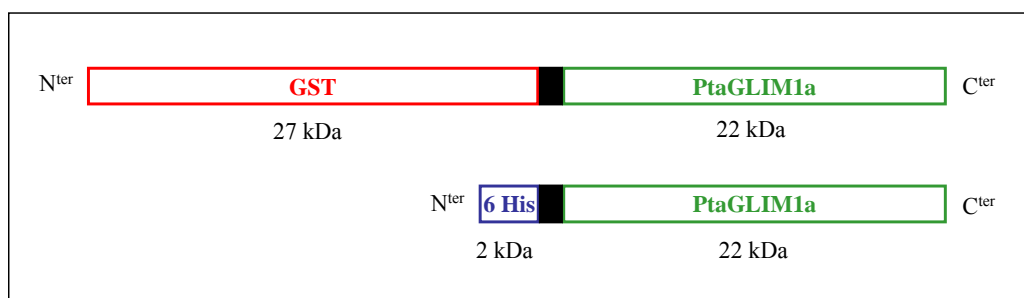
### Partie 3. La protéine PtaGLIM1a

GWLIM1F « 5'-CACCTGGTGCCGCGCGGCAGCATGGCAGCCTTTTCAGGA-3' » et antisens GWLIM1R « 5'-TTATGATTCAGCAGCAATCT-3' ». L'amorce sens GWLIM1F contient le site de reconnaissance par la topoisomérase I « 5'-CACCT-3' » et le site de clivage par la thrombine « 5'-CTGGTGCCGCGCGGCAGC-3' » suivie de la séquence gène spécifique de *PtaGLIM1a*. L'amplification PCR a été réalisée en utilisant la « Platinum<sup>®</sup> Pfx DNA polymerase » (Invitrogen<sup>™</sup>) selon les instructions du fabricant. Après purification du produit PCR avec le kit « Montage<sup>™</sup> PCR Centrifugal Devices » (Millipore), celui-ci a été cloné dans le vecteur pENTR<sup>™</sup>/D-TOPO<sup>®</sup> (Invitrogen<sup>™</sup>) (Annexe 7) selon les instructions du fabricant et en diminuant tous les volumes de moitié. Les clones de bactéries JM109 transformées avec la construction pENTR-PtaGLIM1a ont été sélectionnés sur milieu LB en présence de kanamycine (Kan) à 50 µg/ml, puis criblés par PCR avec les amorces GWLIM1F et GWLIM1R. L'ADN plasmidique des clones positifs en PCR a été extrait avec le kit « GenElute<sup>™</sup> Plasmid Miniprep Kit » (SIGMA<sup>®</sup>) en suivant les instructions du fabricant puis vérifié par digestion avec des enzymes de restriction.

#### *1.6. Recombinaison du gène PtaGLIM1a dans les vecteurs de destination pDEST<sup>™</sup>15 et pDEST<sup>™</sup>17 avec le système Gateway<sup>®</sup>*

Après l'obtention d'un vecteur d'entrée pENTR-PtaGLIM1a par le clonage de la séquence codante *PtaGLIM1a* dans le vecteur pENTR<sup>™</sup>/D-TOPO<sup>®</sup>, la réaction de recombinaison entre ce vecteur d'entrée et les vecteurs de destination pDEST<sup>™</sup>15 et pDEST<sup>™</sup>17 (Annexe 8) a été réalisée avec l'enzyme LR Clonase<sup>™</sup> en suivant les instructions du fabricant et en divisant tous les volumes par quatre. Après transformation de bactéries compétentes JM109, celles ayant intégré les vecteurs d'expression pDEST17-PtaGLIM1a et pDEST15-PtaGLIM1a ont été sélectionnées sur milieu LB contenant de l'ampicilline (Amp) à 100 mg/L. L'ADN plasmidique des clones transformés a été extrait, puis un crible des vecteurs a été réalisé par digestion avec des enzymes de restriction, et par séquençage pour s'assurer qu'aucune mutation n'a été introduite dans la séquence codante des gènes *PtaGLIM1a* et *GST*, ainsi qu'au niveau de l'étiquette 6His et du site de clivage par la thrombine. Ainsi les deux constructions codent bien pour des protéines fusions GST-PtaGLIM1a et 6His-PtaGLIM1a entières (Fig. 55).

### Partie 3. La protéine PtaGLIM1a



**Figure 55.** Représentation schématique des protéines recombinantes GST-PtaGLIM1a et 6His-PtaGLIM1a. Les vecteurs d'expression pDEST15-GLIM1a et pDEST17-GLIM1a construits avec le système Gateway® produisent respectivement les protéines recombinantes GST-PtaGLIM1a et 6His-PtaGLIM1a. Les étiquettes GST et 6His ont des masses moléculaires respectives de 27 et 2 kDa. La protéine PtaGLIM1a a une masse moléculaire estimée d'environ 22 kDa. Un site de clivage par la thrombine indiqué par un rectangle noir, a été introduit entre les étiquettes et la protéine PtaGLIM1a.

#### 1.7. Production de la protéine recombinante GST-PtaGLIM1a

##### 1.7.1. Mise au point des conditions de culture des bactéries pour la production de protéine recombinante GST-PtaGLIM1a

Ces expériences ont été réalisées en collaboration avec Hélène Bénédicti du Centre de Biophysique Moléculaire (CBM) du CNRS d'Orléans. Des précultures de 3 ml dans du milieu LB contenant des antibiotiques appropriés sont ensemencées à partir de colonies de souches de bactéries Rosetta™(DE3) ou Rosetta™ 2(DE3) pLysS transformées avec la construction pDEST15-PtaGLIM1a. Les souches Rosetta™(DE3) et Rosetta™ 2(DE3) pLysS portant respectivement les vecteurs pRARE et pLysSRARE2 sont résistantes au chloramphénicol (Cam) à 34 µg/ml et l'événement de transformation par la construction pDEST15-PtaGLIM1a est sélectionné sur Amp 100 µg/ml. Du sulfate de zinc (ZnSO<sub>4</sub>) ou de l'éthanol à des concentrations finales respectives de 50 µM et de 2 % ont été ajoutés dans les précultures et cultures de bactéries pour tester leurs effets sur la production de protéine recombinante. Les précultures sont effectuées soit à 37°C pendant 12 h, soit à 25°C ou 19°C pendant 24 h, et servent à ensemencer des cultures de bactéries à 0,3 unité de DO<sub>600nm</sub>/ml dans 3 ml de milieu LB avec antibiotiques. L'induction de la production de protéine recombinante est effectuée lorsque la DO<sub>600nm</sub> atteint 0,8 unité/ml, en ajoutant de l'isopropyl β-D-1-thiogalactopyranoside (IPTG) à une concentration finale de 0,1 mM. Les bactéries sont incubées pendant 2h30 à 37°C ou toute la nuit à 25°C ou 19°C, avant la lyse des cellules et l'extraction des protéines.



### Partie 3. La protéine PtaGLIM1a

#### 1.7.2. Mise au point de la solubilisation des protéines recombinantes GST-PtaGLIM1a

Pour les échantillons de protéines totales, un volume de culture correspondant à une  $DO_{600nm}$  de 0,5 est centrifugé 3 min à 9100 g. Le culot de bactéries est repris dans 33,3  $\mu$ l d'eau et 16,6  $\mu$ l de Laemmli 3X [SDS 5 % ; glycérol 30 % ; Tris-HCl 375 mM à pH 6.8,  $\beta$ -mercaptoéthanol 5 % (v/v), bleu de bromophénol] (Laemmli, 1970).

Pour le fractionnement en protéines solubles et insolubles (agrégées ou sous forme de corps d'inclusion), un volume de culture de bactérie équivalent à une  $DO_{600nm}$  de 1 est centrifugé 3 min à 9100 g. Le culot de bactérie est resuspendu dans 100  $\mu$ l de tampon de lyse des bactéries [tampon phosphate de sodium (NaPi) 50 mM à pH 8 ; tampon Tris-HCl 10 mM à pH 8 ; NaCl 250 mM]. Ensuite, sont successivement ajoutés du lysozyme (SIGMA<sup>®</sup>), de la DNase I (Roche Applied Science) et du  $MgCl_2$ , à des concentrations finales respectives de 200  $\mu$ g/ml, 30  $\mu$ g/ml, et 10 mM. En fonction de la dégradation des protéines recombinantes lors de l'extraction des protéines bactériennes, nous avons ajouté 1  $\mu$ l d'un mélange d'inhibiteurs de protéases (IP) de bactéries [AEBSF 23 mM ; EDTA 100 mM ; Bestatin 2 mM ; Pepstatin 0,3 mM ; E-64 0,3 mM] (SIGMA<sup>®</sup>), et du PMSF à une concentration finale de 1 mM. Pour améliorer la solubilité des protéines recombinantes, des détergents tels que du Triton X100 ou du dodécylmaltoside (DDM) sont ajoutés à des concentrations finales respectives de 0,1 % et 0,4 % (v/v). Après agitation, les bactéries sont lysées par le lysozyme pendant 10 min dans la glace. Les bactéries sont ensuite éclatées par 4 à 6 cycles de gel/dégel dans un bain d'alcool froid (-30°C) ou de l'azote liquide, et un bain d'eau à température ambiante. Le mélange est ensuite incubé 15 min à température ambiante pour permettre la dégradation de l'ADN par la DNase I. Les protéines solubles et le cytoplasme sont séparés des protéines insolubles, du peptidoglycane et des membranes par 15 min de centrifugation à 16000 g à 4°C. A 100  $\mu$ l de surnageant est ajouté 33,3  $\mu$ l de Laemmli 3X, puis le mélange est séché à 95°C pour obtenir un volume final de 100  $\mu$ l. Le culot est repris dans 33,3  $\mu$ l de Laemmli 3X et 66,6  $\mu$ l d'eau. Avant la migration des protéines sur gel SDS-PAGE, les échantillons sont chauffés pendant 5 min à 95°C, puis centrifugés pendant 5 min à 16000 g pour culotter les éventuels débris cellulaires, ainsi que les protéines restées insolubles.

### Partie 3. La protéine PtaGLIM1a

#### 1.7.3. Production et purification sur de la protéine recombinante GST-PtaGLIM1a

Des clones de bactéries Rosetta™(DE3) transformées avec la construction pDEST15-PtaGLIM1a sont ensemencées dans des précultures de 3 ml puis de 50 ml à 25°C dans du milieu LB liquide avec sélection sur Amp (100 µg/ml) et Cam (34 µg/ml). Ces précultures permettent d'ensemencer quatre cultures de bactéries de 125 ml à 0.3 unité de DO<sub>600nm</sub>/ml. Lorsque la DO<sub>600nm</sub> atteint 0,8 unité/ml, de l'IPTG à concentration finale de 0,1 mM est ajouté pour une production de protéine recombinante durant la nuit. Les bactéries sont centrifugées pendant 5 min à 9100 g, et le culot est repris dans 9 ml de tampon de lyse des bactéries contenant du DDM à 0,1%, du lysozyme à 200 µg/ml, des inhibiteurs de protéases dilués au 1/100<sup>e</sup>, de la DNase I à 5 µg/ml, du MgCl<sub>2</sub> à 10 mM, et du PMSF à 1 mM. Le mélange est incubé 20 min dans la glace, puis 4 cycles de gel/dégel (azote liquide) sont appliqués pour l'éclatement des bactéries. Du DDM à une concentration finale de 0,4 % est ajouté puis le mélange est incubé 30 min dans la glace pour la solubilisation de la protéine GST-PtaGLIM1a. Les protéines insolubles sont culottées par une ultracentrifugation à 100000 g pendant 30 min à 4°C, et les protéines solubles du surnageant sont filtrées à travers un filtre de 45 µm (MILLIPORE®). Du PMSF et du DTT sont ajoutés aux protéines filtrées pour des concentrations finales de 1 mM.

Au préalable, 0,75 ml de gel de « Glutathione Sepharose 4B » (Amersham Biosciences®) est équilibré par quatre lavages avec 3 ml de tampon de lyse des bactéries en appliquant la méthode batch avec des centrifugations à 500 g pendant 5 min et à 4°C. Les protéines GST-PtaGLIM1a sont fixées au gel de Glutathione Sepharose 4B par 30 min d'incubation à 4°C. Après une centrifugation à 500 g pendant 15 min à 4°C pour éliminer les protéines bactériennes non fixées, le gel est ensuite transféré dans une colonne à filtration. La colonne est lavée avec 15 ml de tampon de lavage de la colonne [tampon phosphate de sodium 50 mM à pH 8 ; NaCl 150 mM, KCl 50 mM ; glycérol 5 % ; PMSF 1 mM ; EDTA 1 mM ; Triton X100 0,1 % ; acétate de sodium pH 7 10 mM] pour éliminer les protéines non fixées au gel. Deux autres lavages avec 15 ml de tampon de lavage contenant du Triton X100 1 % puis du NaCl 300 mM sont effectués pour éliminer les protéines bactériennes fixées de manière non spécifique. Avant élution de la protéine GST-PtaLIM1a, la colonne est équilibrée avec du tampon de transition [Tris-HCl 50 mM à pH 8 ; NaCl 150 mM]. La protéine GST-PtaLIM1a est éluee avec trois fois 2 ml de tampon d'élution [Tris-HCl 10 mM à pH 8 ; glutathion réduit 10 mM]. Le glutathion libre resté dans le gel est éliminé par quatre lavages

### Partie 3. La protéine PtaGLIM1a

en batch avec 50 ml le tampon [Tris HCl 0,1 M à pH 8,5 ; NaCl 0,5 M]. La colonne est ensuite régénérée selon les instructions du fabricant, puis rééquilibrée avec du tampon de lyse des bactéries.

#### 1.7.4. Clivage à la thrombine de la protéine GST-PtaGLIM1a et capture de la thrombine par la streptavidine

Le glutathion libre présent dans l'éluat est éliminé par dialyse à 4°C toute la nuit dans du tampon de clivage à la thrombine [20 mM Tris-HCl pH 8,5, 100 mM NaCl, 330 µM CaCl<sub>2</sub>]. Cette dialyse est réalisée avec une membrane de dialyse en cellulose (SPECTRA/POR<sup>®</sup>) laissant passer les molécules ayant une masse moléculaire inférieure à 12000-14000 Da. Les protéines dialysées sont centrifugées à 16000 g pendant 10 min à 4°C pour éliminer les éventuelles protéines précipitées. Pour le clivage de la protéine GST-PtaGLIM1a par la thrombine, le kit « Thrombin Cleavage Capture Kit » (Novagen<sup>®</sup>) a été utilisé selon les instructions du fabricant. La thrombine a été utilisée à une concentration de 1,5 unité/ml et pour définir les conditions optimales de clivage de la GST, une cinétique de la réaction enzymatique entre 1 et 16 heures a été testée, sous agitation douce, et à température ambiante ou à 4°C. Après clivage de la GST à 4°C pendant 1h30, la thrombine est ensuite éliminée par capture avec de la streptavidine couplée à des billes d'agarose, selon les instructions du fabricant. La réaction est effectuée à température ambiante au minimum 30 min sous agitation douce et à raison d'un ratio de 16 µl de gel d'agarose sec par unité de thrombine. Avant l'élimination de la GST libre sur la colonne Glutathione Sepharose 4B, du DDM à une concentration finale de 0,4 % est ajouté pour conserver la solubilité de la protéine PtaGLIM1a.

#### 1.7.5. Elimination de la GST sur colonne Glutathione Sepharose 4B :

Avant la fixation de la GST sur le gel de Glutathione Sepharose 4B, le pH de la solution de protéines est ajusté à 7,5 avec du tampon Tris-HCl 1 M à pH 7,5 ; puis du DTT est ajouté pour une concentration finale de 1 mM. Les protéines sont incubées à 4°C pendant 30 min en présence du gel de Glutathione Sepharose 4B pour la capture de la GST. Après une centrifugation à 500 g pendant 15 min à 4°C, le surnageant contenant la protéine PtaGLIM1a purifiée est récupéré et pour conserver la protéine à -20°C, du glycérol est ajouté à une

### Partie 3. La protéine PtaGLIM1a

concentration finale de 20 %. Après lavage de la colonne, la GST fixée sur le gel de Glutathione Sepharose 4B est éliminée par des lavages avec le tampon d'élution. Enfin, la colonne est régénérée selon les instructions du fabricant.

#### 1.7.6. Test d'élimination des protéines contaminantes par tamisage moléculaire sur une colonne de gel filtration.

Afin d'éliminer les protéines contaminantes de poids moléculaire de 58 kDa et 72 kDa, nous avons utilisé, avec l'aide de Franck Coste du CBM d'Orléans, une colonne « Superdex 75 » (Pharmacia) de 50cm de long. Pour cela, la colonne a été équilibrée durant la nuit avec le tampon d'équilibration froid [Tris-HCl 50 mM à pH 7,4 ; NaCl 250 mM ; MgSO<sub>4</sub> 10 mM ; Triton X100 0,1 % ; glycérol 5 %] filtré sur 0,2 µm et dégazé pendant 15 min. Les protéines ont été lavées, équilibrées et concentrées à 4°C dans du tampon d'équilibration en utilisant des filtres à centrifugation « Ultrafree Biomax » (MILLIPORE®) laissant passer les molécules d'un poids inférieur à 10 kDa. De l'ATP à une concentration finale de 2 mM est ajouté aux protéines, puis une incubation à 37°C pendant 10 min est appliquée pour rompre l'éventuelle association de la protéine PtaGLIM1a avec la protéine contaminante dnaK d'*E. coli*. La solution de protéine est filtrée à 0,22 µm avant injection sur la colonne « Superdex 75 ». A la sortie de la colonne, l'absorbance de la solution ainsi tamisée est mesurée à 280 nm pour vérifier la présence de protéines lors de la collecte des fractions de 1 ml. Ces fractions seront ensuite concentrées puis analysées en gel SDS-PAGE.

#### 1.7.7. Test d'élimination des protéines chaperonnes GroEL

Avant l'élution de la protéine GST-PtaGLIM1a par le glutathion, la colonne de gel Glutathione Sepharose 4B est équilibrée avec la solution EB1 [tampon TEA-HCl 50 mM à pH 7,5 ; NaCl 150 mM ; KCl 50 mM ; MgCl<sub>2</sub> 20 mM]. Pour l'élimination de GroEL, la colonne est lavée avec la solution EB2 (solution EB1 contenant de l'ATP à 10 mM), puis dans la solution EB2 contenant du Triton X100 1% et/ou une concentration en NaCl de 300 mM. La colonne est ensuite équilibrée avec le tampon de transition, avant l'élution de la protéine GST-PtaGLIM1a.

## Partie 3. La protéine PtaGLIM1a

### 1.8. Production et purification de la protéine recombinante 6His-PtaGLIM1a

#### 1.8.1. Test de la production et de la solubilisation de la protéine 6His-PtaGLIM1a dans différentes bactéries

Pour ces expériences en collaboration avec Hélène Bénédicti, j'ai également été conseillé par Bertrand Castaing, et Alain Roussel (CBM Orléans) qui nous a permis de compléter notre collection de bactéries. Ainsi, les souches de bactérie BL21(DE3), BL21(DE3) pLysS, Rosetta™(DE3), Rosetta™ 2(DE3) pLysS, Tuner(DE3) pLysS, Origami(DE3), Origami(DE3) pLysS, C41 et C43 ont été transformées par choc thermique avec la construction pDEST17-PtaGLIM1a. La souche Origami est sélectionnée avec de la Kan à 15 µg/ml et de la Tétracycline (Tet) à 12,5 µg/ml. La construction pDEST17-PtaGLIM1a et le vecteur pLysS sont respectivement sélectionnés avec de l'Amp à 100 µg/ml et du Cam à 34 µg/ml. Pour la protéine 6His-PtaGLIM1a, les conditions de culture des bactéries et de production de protéines recombinantes après induction à l'IPTG sont les mêmes que pour la protéine GST-PtaGLIM1a, exceptés des temps d'induction de 4 h à 37°C ou d'une nuit à 24°C. De plus, je n'ai pas ajouté de zinc ou de l'éthanol dans les précultures. De même, pour l'extraction et la solubilisation de la protéine 6His-PtaGLIM1a puis l'analyse des fractions soluble et insoluble, j'ai utilisé la méthode employée pour la protéine GST-PtaGLIM1a, sauf que le tampon de lyse des bactéries ne contient pas de détergents, ni d'inhibiteurs de protéases et de PMSF.

#### 1.8.2. Purification de la protéine recombinante 6His-PtaGLIM1a sous forme de corps d'inclusion puis par affinité sur une résine de Cobalt

Les bactéries de la souche BL21(DE3) pLysS transformées avec la construction pDEST17-PtaGLIM1a sont utilisées pour ensemercer des précultures de 3 ml puis 50 ml à 37°C dans du milieu LB liquide avec sélection sur Amp (100 µg/ml) et Cam (34 µg/ml). Ces précultures permettent d'ensemencer quatre cultures de bactéries de 500 ml à 0.3 unité de DO<sub>600nm</sub>/ml. Lorsque la DO<sub>600nm</sub> de la culture atteint 0,8 unité/ml, de l'IPTG à concentration finale de 0,1 mM est ajouté pour une production de protéine recombinante à 37°C pendant 4 h. Les bactéries sont centrifugées à 9100 g pendant 5 min.

Le culot de bactéries est repris dans 30 ml de tampon CI de lyse pour les corps d'inclusion [tampon Tris-HCl 10 mM à pH 8 ; EDTA 1,5 mM ; saccharose 25 % ; DTT 10

### Partie 3. La protéine PtaGLIM1a

mM] puis du lysozyme, de la DNase I et du  $MgCl_2$  sont successivement ajoutés à des concentrations finales respectives de 200  $\mu g/ml$ , 5  $\mu g/ml$ , et 10 mM. Le mélange est incubé 30 min dans la glace pour l'action du lysozyme, puis 4 cycles de gel/dégel sont appliqués pour éclater les cellules. Le mélange est ensuite incubé 15 min à température ambiante pour l'action de la DNase I. Les corps d'inclusion contenant la protéine recombinante 6His-PtaGLIM1a sont culottés par 15 min d'ultracentrifugation à 80000 g et à 4°C. Les culots de corps d'inclusion sont ensuite lavés quatre fois avec du tampon CI contenant du Triton X100 à une concentration finale de 0,5 %. Pendant les lavages, les culots sont repris à la spatule et centrifugés à 80000 g pendant 15 min à 4°C. Un dernier lavage est effectué avec du tampon CI sans Triton X100.

Avant purification sur une résine de cobalt « BD TALON™ Metal Affinity Resin » (BD Biosciences Clontech), 20 mg de protéine recombinante 6His-PtaGLIM1a sous forme de corps d'inclusion sont lavés deux fois en batch avec 50 ml de tampon phosphate salin [NaPi 50 mM à pH 7 ; NaCl 300 mM] pour éliminer l'EDTA, le Tris-HCl et le DTT incompatibles avec la résine de cobalt. Entre chaque lavage, les corps d'inclusion sont centrifugés à 21000 g pendant 10 min à 4°C. Le culot de corps d'inclusion est solubilisé par une heure d'incubation à 4°C dans 30 ml de tampon SL [NaPi 50 mM à pH 7 ; NaCl 300 mM ; urée 8 M]. Les protéines insolubles dans l'urée sont centrifugées à 21000 g pendant 15 min à 4°C, et le surnageant de protéines solubles est filtré à travers un filtre de 0,45  $\mu m$ . La résine de cobalt est équilibrée par trois lavages en batch dans 20 ml de tampon SL avec des centrifugations à 1500 g pendant 5 min à 4°C. Pour la fixation de la protéine 6His-PtaGLIM1a à la résine de cobalt, le filtrat de protéines est incubé avec la résine pendant 1 h à 4°C sous agitation. La résine de cobalt est ensuite lavée trois fois en batch pendant 10 min dans 50 ml de tampon SL avec des centrifugations à 1500 g. La protéine 6His-PtaGLIM1a est éluée avec deux fois 5 ml de tampon d'éluion [tampon NaPi 50 mM à pH 7 ; NaCl 300 mM ; arginine 0,8 M ; imidazole 400 mM ; glycérol 5%], pendant 15 min sous agitation et à 4°C. La résine de cobalt est régénérée par un lavage dans 50 ml de tampon [MES 20 mM à pH 5 ; NaCl 0,1 M], puis est conservée dans de l'eau distillée. Les protéines 6His-PtaGLIM1a éluées sont ensuite successivement dialysées pendant au minimum 4 h dans 1 litre de trois solutions de dialyse différentes : une solution de dialyse D1 [tampon Tris-HCl 20 mM à pH 7,4 ; NaCl 500 mM ; glycérol 5% ;  $ZnSO_4$  1 mM ; arginine 0,5 M], une solution de dialyse D2 [tampon Tris-HCl 20 mM à pH 7,4 ; NaCl 500 mM ; glycérol 15% ;  $ZnSO_4$  1 mM ; arginine 0,2 M] et une solution de dialyse D3 [20 mM tampon Tris-HCl pH 7,4 ; NaCl 500 mM ;  $ZnSO_4$  1 mM ; glycérol 10%].

### Partie 3. La protéine PtaGLIM1a

La solubilité des protéines 6His-PtaGLIM1a dialysées dans la solution D3 a été testée par filtration à travers un filtre de 0,22 µm, puis par concentration avec des unités à filtration de 500 µl « Nanosep<sup>®</sup> Centrifugal Devices 3 K », (Pall<sup>®</sup> Life Sciences) laissant passer les molécules inférieures à 3 kD. Les protéines sont concentrées avec ces unités à filtration par centrifugation à 12000 g pendant environ 15 min à 4°C. Les protéines insolubles et solubles sont séparées par centrifugation à 16000 g pendant 15 min, puis analysées sur gel SDS-PAGE et colorées au bleu de Coomassie. La concentration des protéines solubles a également été estimée par des mesures de l'absorbance à 280 nm avec un « NanoDrop<sup>®</sup> ND-1000 Spectrophotometer » (Nanodrop Technologies Inc).

#### *1.9. Production d'anticorps polyclonaux anti-PtaGLIM1a :*

Après lavage des corps d'inclusion et purification sur colonne d'affinité, la pureté de la protéine 6His-PtaGLIM1a est suffisante (environ 95 %) pour la production d'anticorps spécifiques. Après concentration jusqu'à 0,5 mg/ml de protéine soluble, et sachant que beaucoup de protéines ont précipité, 4 ml de protéines 6His-PtaGLIM1a (2 mg) dans le tampon de dialyse D3 ont été envoyées à la société AGRO-BIO (La Ferté St Aubin, France) pour la production d'anticorps à partir de deux lapins. Ces lapins ont été immunisés en 77 jours par cinq injections à 0, 14, 28, 42 et 56 jours. Les sérums de lapins ont été prélevés à 0 (sérum pré-immun), 49, 63 et 77 jours.

D'autre part le peptide P-GLIM1a ayant la séquence suivante « H-CNEREKNPVVSITAVE-NH<sub>2</sub> » a été synthétisé. Il se situe à la partie C-terminale de la protéine PtaGLIM1a. C'est la région la plus variable par rapport aux autres membres de la famille de protéines LIM chez le peuplier (Fig. 56B). Il a été commandé à la société AGRO-BIO qui a réalisé la synthèse de 5 mg de peptide et son couplage à 10 mg de la protéine porteuse KLH. Puis deux lapins ont été immunisés avec le peptide P-GLIM1a couplé à la KLH en 77 jours par cinq injections à 0, 14, 28, 42 et 56 jours. Les sérums de lapins ont été prélevés à 0 (sérum pré-immun), 49, 63 et 77 jours.

Par ailleurs, nous avons reçu de la part de Mr Michel Petitprez de l'ENSA de Toulouse un anticorps A16 (lapins 1 et 2) purifié et dirigé contre le peptide P16 « CDQLFKKTGSLDKSFE » situé dans la région inter-LIM de la protéine HaWLIM1 (Brière *et al.*, 2003). Cette séquence est quasiment identique aux régions inter-LIM des protéines PtaGLIM1a, PtaGLIM1b, PtaWLIM1a, PtaWLIM1b avec une substitution conservée. Elle est

### Partie 3. La protéine PtaGLIM1a

aussi similaire aux régions inter-LIM des protéines Pta $\beta$ LIM1a et Pta $\beta$ LIM1b, mais avec un ou deux acides aminés très différents (Fig. 56A).

**A**

```

PtaGLIM1a      YDQLFKRRTGSLDKSFEoGTPKIVKPEKPVNDNEN--ASKVSNLFAQTREK
PtaGLIM1b      YDQLFKRRTGSLDKSFEoGTPKIAKPEKPVNDNEN--ASKVSNLFAQTREK
PtaWLIM1a      FDQLFKRRTGSLDKSFEoGTPKIVKPEKPVNDGKPVSTKVSTMFAGTREK
PtaWLIM1b      FDQLFKRRTGSLDKSFEoGTPKIVKPEKPVNDGKPISTKVSTMFAGTRDK
Pt•LIM1a       FDQLLKMTGSLDKSFEoKGTSKTVRVDRSADQVQS-NSKVSSMFAQTQEK
Pt•LIM1b       FDQLFKMTGSLDKSFEoGTPKTVRGDRSADQVLS-NSKVSSMFAQTQEK
PtaWLIM2a      FDQLFKETGNFNKNFQSPAKTAEKLTPELTRSP--SKAAGMFSGTQEK
PtaWLIM2b      FEQLFKETGNFNKNFQSPAKSAEKLNPELTRSP--SKAASMFSGTQEK
PtaPLIM2a      FEQLFKEGGDFSRNFQ---KGKPERTHDLSRIP--SKLSSVFCGTQDK
PtaPLIM2b      FEQLFKEGGDFSKNFQ---KGKPERTHELIRTP--SKLSSVFCGTQDK
Pt•LIM2a       YDQLFKETGNFTKKFQ-----PCEIRFSTKAP--SKLSSMFSGTQDK
Pt•LIM2b       YDQLFKETGNFSKKLQ-----SCEIPLFTKAP--SKLSSMFSGTQDK
::**:*   *.: :.: :   :* : :*.***:*

```

**B**

```

PtaGLIM1a      HIQLFKEKGNYSQLENEREKNPV-oVSITAVE-----IAAES-----
PtaGLIM1b      HIQLFKEKGNYSQLENEREKNPVTVNI TAVE-----IAAES-----
PtaWLIM1a      HNQLIKEKGNLSQLEGDIEKDSMNKNTNGRE-----VAAES-----
PtaWLIM1b      HNQLIKEKGNLSQLEGDVEKDSMNDKNTNGRE-----VSAES-----
Pt•LIM1a       HNQLFKQKGNFSQLDKHEHVTPVKETAA-----AE-----
Pt•LIM1b       HNQLFKEKGNFSQFGKHEHLTPVDETAT-----AE-----
PtaWLIM2a      FSQLFKEKGSYNHLIKSASMKRAAAP-----VPEA-----
PtaWLIM2b      FSQLFKEKGSYNHLIKSATMKRAAAS-----VPEA-----
PtaPLIM2a      FAQLFMEKGTYSHVLASAAHKRSNSTTP-----ELAGSNPEEGAAAEKSEEQS
PtaPLIM2b      FAQLFMEKGTYSHVLAGATHKRSTSTPPP-----ELAESNTEEAAD-KSEEQS
Pt•LIM2a       FAQLFKQKGSYSYLTKTATMKKNAVNLPEEKSEAAENNETVPEANADANSDLAIDKRHHNS
Pt•LIM2b       FSQLFKQKGSYSYLTKTSTMKKNVNSPEEKSEAEQNHLTVPEA-----SSDLAIAHENVQN
.  ** : ** . . .

```

**Figure 56.** Localisation des peptides de synthèse ayant servi à la production d'anticorps A16 et anti-P-GLIM1a  
**(A)** Alignement du domaine inter-LIM des protéines LIM de peuplier par le logiciel ClustalW. Le peptide P16 « CDQLFKKTGSLDKSFE » désigné à partir de la protéine HaWLIM1 (Brière et *al.*, 2003) est localisé au niveau de la région inter-LIM des protéines du groupe LIM1 de peuplier. En rouge sont indiqués les résidus identiques et en orange sont indiqués les résidus similaires.  
**(B)** Alignement du domaine C-terminal des protéines LIM de peuplier par le logiciel ClustalW. Le peptide de synthèse P-GLIM1a « CNEREKNPVVSITAVE » indiqué en rouge est spécifique de la protéine PtaGLIM1a au niveau de sa région C-terminale.

#### 1.10. Extraction de protéines totales de peuplier

Les protéines totales de peuplier ont été préparées à partir des mêmes prélèvements de tissus ayant servi à l'analyse de l'expression des gènes *PtaLIM* par RT-PCR semi quantitative (Cf. Matériel et méthodes de la partie 2). Ainsi, nous avons une répétition biologique pour les tissus provenant d'un arbre âgé de quatre ans et trois répétitions biologiques pour les tissus provenant d'arbres âgés de trois mois et cultivés en serre. Pour ces derniers, les échantillons broyés de trois arbres indépendants inclinés ont été mélangés pour la préparation de protéines totales. Quel que soit le tissu analysé, les protéines ont été extraites avec une méthode de précipitation des protéines au TCA et à l'acétone (Damerval *et al.*, 1986). 200 à 300 mg de tissus broyés sont précipités pendant une nuit à -20°C dans une solution de précipitation des protéines froide [acétone, TCA 10 %,  $\beta$ -mercaptoéthanol 0,07 %]. Les protéines sont centrifugées pendant 30 min à 16000 g à 4°C, puis le culot de protéine est lavé au minimum



### Partie 3. La protéine PtaGLIM1a

trois fois dans une solution de rinçage de protéines froide [acétone,  $\beta$ -mercaptoéthanol 0,07 %], afin d'éliminer les phénol et la chlorophylle. Lorsque le culot a un aspect blanchâtre, il est directement repris, pendant 1 heure à température ambiante et sous agitation, dans un tampon de charge des protéines [Tris-HCl 80 mM à pH 6,8 ; DTT 0,1 M ; SDS 2 % ; urée 4 M ; glycérol 10 % ; bleu de bromophénol] (Baltz *et al.*, 1999). Les protéines insolubles et les débris cellulaires sont sédimentés par centrifugation pendant 15 min à 16000 g et le surnageant contenant les protéines solubles est conservé à -20°C jusqu'à utilisation.

#### 1.11. *Dosage de la quantité de protéine par la méthode de Bradford (Bradford, 1976)*

Une aliquote de protéine dans le tampon de charge est prélevée puis précipitée pendant 15 min dans la glace avec du TCA à une concentration finale de 15 %. Les protéines précipitées sont culottées par une centrifugation de 10 min à 16000 g à 4°C. Le culot de protéine est lavé deux fois avec de l'acétone froid, pour éliminer le TCA, puis le culot est séché au speed vac avant d'être resuspendu dans une solution de soude (NaOH) à une concentration finale de 0,1 N. Une gamme d'albumine de sérum bovin (BSA) de 0 à 0,3  $\mu\text{g}/\mu\text{l}$  et réalisée dans d'une solution de NaOH à 0,1 N. Trois répétitions de 10  $\mu\text{l}$  de cette gamme sont déposées dans une plaque de lecture de 96 puits et en parallèle, 1 à 10  $\mu\text{l}$  d'échantillons de protéines (trois répétitions) est dosé. Le réactif de Bradford « Bio-Rad Protein Assay » (BIORAD®) constitué de bleu de Coomassie G-250 est dilué cinq fois, puis un volume de 250  $\mu\text{l}$  est ajouté aux échantillons. Après une incubation de 30 min sous agitation, l'absorbance à 595 nm des échantillons de la plaque est lue par un spectrophotomètre « Multiscan spectrum » (Thermo electron corporation).

#### 1.12. *Migration des protéines en condition dénaturante en gel SDS-PAGE (Laemmli, 1970)*

La migration des protéines s'effectue avec l'appareillage «Miniprotean II » (BIORAD®), dans un tampon de migration [Tris-base 25 mM ; glycine 192 mM à pH 7,4 ; SDS 0,1 %] à 200 V pendant 1 h environ. Les protéines migrent d'abord à travers un « stacking gel » [Tris-HCl 125 mM à pH 6,8 ; acrylamide/bis-acrylamide (37,5/2,5) 4 % ; SDS 0,1 % ; APS 0,1 % (w/v), Temed 0,1 % (v/v)] pour concentrer les protéines avant leurs

### Partie 3. La protéine PtaGLIM1a

séparations à travers un gel de migration ou « running gel » [Tris-HCl 375 mM à pH 8,8 ; acrylamide/bis-acrylamide (37,5/2,5) à 10 ou 12 % ; SDS 0,1 % ; APS 0,1 % (w/v) ; Temed 0,1 % (v/v)].

Pour les tests de production des protéines recombinantes GST-PtaGLIM1a et 6His-PtaGLIM1a, 10 µl et 15 µl de protéines ont été respectivement chargées pour le Western-blot et la coloration des protéines au bleu de Coomassie. Pour les échantillons de protéines de peuplier, 10 ou 20 µg de protéines totales sont chargées sur gel. Pour les analyses en spectrométrie de masse MS/MS des protéines contaminantes, les échantillons de protéines ont d'abord été réduits avec du Tris [2-carboxyéthyl] phosphine hydrochloride (TCEP-HCl) un agent réducteur fort, et alkylés avec de la iodoacétamine (IAA) pour bloquer la formation de ponts disulfures. De plus, toutes les précautions ont été prises pour éviter une contamination avec la kératine de la peau ou des cheveux lors de la préparation des échantillons et de la migration SDS-PAGE. Les bandes ont été découpées au scalpel et gardées à 4°C avant analyse en spectrométrie de masse MS/MS, effectuée par Maya Belghazi du « Service de spectrométrie de masse pour la protéomique » (UMR 6175, INRA de Tours).

#### 1.13. *Coloration au bleu de Coomassie :*

Les protéines sont colorées par 30 minutes d'incubation dans une solution de bleu de Coomassie [méthanol 45 % (v/v) ; acide acétique 10 % (v/v) ; bleu de coomassie « Coomassie Brilliant Blue R-250 » (SIGMA<sup>®</sup>) 0,25 % (w/v)]. Les gels sont ensuite lavés deux fois dans une solution de décoloration [éthanol/H<sub>2</sub>O/acide acétique (2/8/1 v/v)], puis lavés à l'eau avant le scan des gels.

#### 1.14. *Transfert des protéines sur membrane de nitrocellulose et réalisation de Dot-blot avec le peptide P-GLIM1a*

Les protéines sont transférées sur une membrane de nitrocellulose Hybond<sup>™</sup>-C-Extra Nitrocellulose (Amersham biosciences<sup>®</sup>), dans du tampon de transfert [Tris-base 25 mM, glycine 192 mM, éthanol 20 %], à 100 V pendant 1 h, ou à 40 V pendant la nuit à 4°C, et sous agitation. L'efficacité du transfert des protéines sur la membrane est vérifiée par coloration des protéines avec une solution de rouge ponceau [TCA 3 % (v/v) rouge ponceau 0,2 % (w/v)] pendant 10 min, et rinçage de la membrane avec de l'eau puis du tampon TBS [Tris-

### Partie 3. La protéine PtaGLIM1a

HCl 20 mM à pH 7,5 ; NaCl 150 mM]. Pour la réalisation des Dot-blot avec le peptide P-GLIM1a, celui-ci a été solubilisé dans de l'eau stérile à une concentration finale de 1 mg/ml, puis dilué dans de l'eau au 1/10<sup>e</sup>, 1/100<sup>e</sup>, et 1/1000<sup>e</sup>. 0,5 µl de ces différentes dilutions a été déposé sur la membrane de nitrocellulose et séché à l'air avant la saturation des membranes.

#### 1.15. Western blot

Pour la réalisation de Western blot par la méthode de colorimétrie, les membranes sont d'abord saturées pendant au minimum 30 minutes avec une solution de TBS contenant du lait écrémé à 5 %. L'hybridation des membranes avec les anticorps primaires s'effectue pendant 1 h à température ambiante ou la nuit à 4°C. Les anticorps monoclonaux anti-6His « His Tag<sup>TM</sup> Monoclonal Antibody » (Novagen<sup>®</sup>) et les anticorps monoclonaux anti-GST « GST Tag<sup>TM</sup> Monoclonal Antibody » (Novagen<sup>®</sup>) sont respectivement dilués au 1/1000<sup>e</sup> et au 1/10000<sup>e</sup> dans 2 ml de tampon TBS contenant du lait écrémé à 5 %. La membrane est ensuite lavée pendant trois fois 5 min avec du tampon TBS contenant du lait écrémé à 5 %. L'hybridation des membranes avec les anticorps secondaires s'effectue pendant 1 h à température ambiante ou la nuit à 4°C. Les anticorps secondaires anti-IgG (H + L) de souris couplés à la phosphatase alcaline « Anti-Mouse IgG (H + L) Alkaline Phosphatase Conjugate » (PROMEGA<sup>®</sup>) à 1 mg/ml sont dilués au 1/7500<sup>e</sup> dans 2 ml de tampon TBS contenant du lait écrémé à 5 %. Les membranes sont ensuite lavées pendant trois fois 5 min dans du tampon TBS puis rincée pendant 5 min avec du tampon de révélation [Tris-HCl 100 mM à pH 9,5 ; NaCl 100 mM ; MgCl<sub>2</sub> 5 mM]. Les membranes sont révélées à l'obscurité et sans agitation avec le « BCIP/NBT Color development substrate » (PROMEGA<sup>®</sup> Madison, WI) par incubation dans 10 ml de tampon de révélation contenant 66 µl de nitro blue tetrazolium (NBT) à 50 mg/ml et 33 µl de 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) à 50 mg/ml. Après révélation, les membranes sont rincées à l'eau et séchées.

Pour la réalisation de Western blot et Dot blot par la méthode de luminescence chimique, nous avons utilisé le kit « ECL Advance<sup>TM</sup> Western blotting detection kit » (Amersham biosciences<sup>®</sup>) en suivant les instructions du fabricant. Cette méthode a été utilisée avec les anticorps primaires A16, anti-6His-PtaGLIM1a et anti-P-GLIM1a pour détecter la protéine recombinante 6His-PtaGLIM1a et les différentes protéines PtaLIM dans les échantillons de protéines totales de peuplier. Les anticorps A16 ont été utilisés selon les recommandations fournies (Brière *et al.*, 2003), et ont été dilués au 1/200<sup>e</sup> dans du tampon

TBS contenant du Tween 20 à 0,1 % (v/v) et du « Blocking reagent » à 2 % (w/v). Deux tests de dilution au 1/1000<sup>e</sup> et au 1/5000<sup>e</sup> ont été appliqués sur les sérums bruts de lapin anti-6His-PtaGLIM1a et anti-P-GLIM1a prélevés à 77 jours. Les anticorps secondaires couplés à la peroxydase de raifort « Horseradish Peroxidase (HPR) » ont été utilisés dilués au 1/50000<sup>e</sup>. Les membranes ont été exposées sous film autoradiographique « KODAK BioMax MR Film » (Kodak<sup>®</sup>) pendant des temps variables de 10 sec à 1 min.

## 2 Résultats

Nous avons décidé de produire une protéine recombinante PtaLIM pour l'objectif principal suivant : l'obtention d'anticorps afin de vérifier l'expression du gène au niveau protéique par des expériences de Western-blot et par la suite d'immuno-histolocalisation. En effet, le bois est un tissu complexe, et il est important de savoir dans quel type cellulaire (fibres, rayons ou vaisseaux) est exprimé la protéine pour en comprendre sa fonction. De plus, sa localisation subcellulaire pourra aussi être examinée. Par ailleurs, l'obtention d'une protéine recombinante soluble est nécessaire pour des études fonctionnelles comme l'analyse de la fixation de la protéine à l'ADN, l'actine ou d'autres protéines partenaires. Pour cela nous avons choisi le gène codant la protéine PtaGLIM1a. En effet, cette protéine appartient au nouveau sous-groupe phylogénétique FLIM1 pour lequel aucune protéine n'a encore été caractérisée. Par ailleurs, c'est surtout le gène dont l'expression est la plus fortement induite dans le bois de tension en particulier au niveau du xylème mature suggérant son implication dans la formation de la paroi secondaire. Pour la production et la purification de protéine recombinante PtaGLIM1a, nous avons utilisé deux étiquettes différentes, Glutathione S Transférase (GST) et 6Histidine (6His), positionnées en N-terminal de la protéine PtaGLIM1a (Fig. 55). Pour cela la partie codante de l'ADNc *PtaGLIM1a* a été clonée dans le vecteur pENTR<sup>™</sup>/D-TOPO<sup>®</sup> grâce à la technologie Gateway. Un site de clivage à la thrombine a été introduit en N-terminal de la protéine PtaGLIM1a pour éliminer l'étiquette après purification. Ce vecteur a servi à insérer le gène par recombinaison dans deux vecteurs de destination différents : pDEST<sup>™</sup>15 pour construire une protéine recombinante fusionnée à la GST et pDEST<sup>™</sup>17 pour construire une protéine recombinante portant l'étiquette 6His. Nous avons décidé dans un premier temps d'utiliser la construction pDEST15-PtaGLIM1a car la GST permet dans de nombreux cas d'exprimer des protéines fusions de manière soluble en solution aqueuse sans agent dénaturant (Smith & Johnson, 1988). De plus, beaucoup d'études

### Partie 3. La protéine PtaGLIM1a

précédentes sur les protéines à domaine LIM de plante (Baltz *et al.*, 1999 ; Kawaoka *et al.*, 2000 ; Mundel *et al.*, 2000 ; Brière *et al.*, 2003) ou CRP animale (Feuerstein *et al.*, 1994 ; Arber & Caroni, 1996) ont utilisé la fusion GST pour la production de protéine recombinantes. Toutefois les résultats obtenus avec la protéine GST-PtaGLIM1a, qui seront décrits ci-dessous, n'ont pas été satisfaisants et m'ont conduit à choisir une stratégie alternative utilisant l'étiquette 6His.

#### 2.1. Production de protéines recombinante GST-PtaGLIM1a :

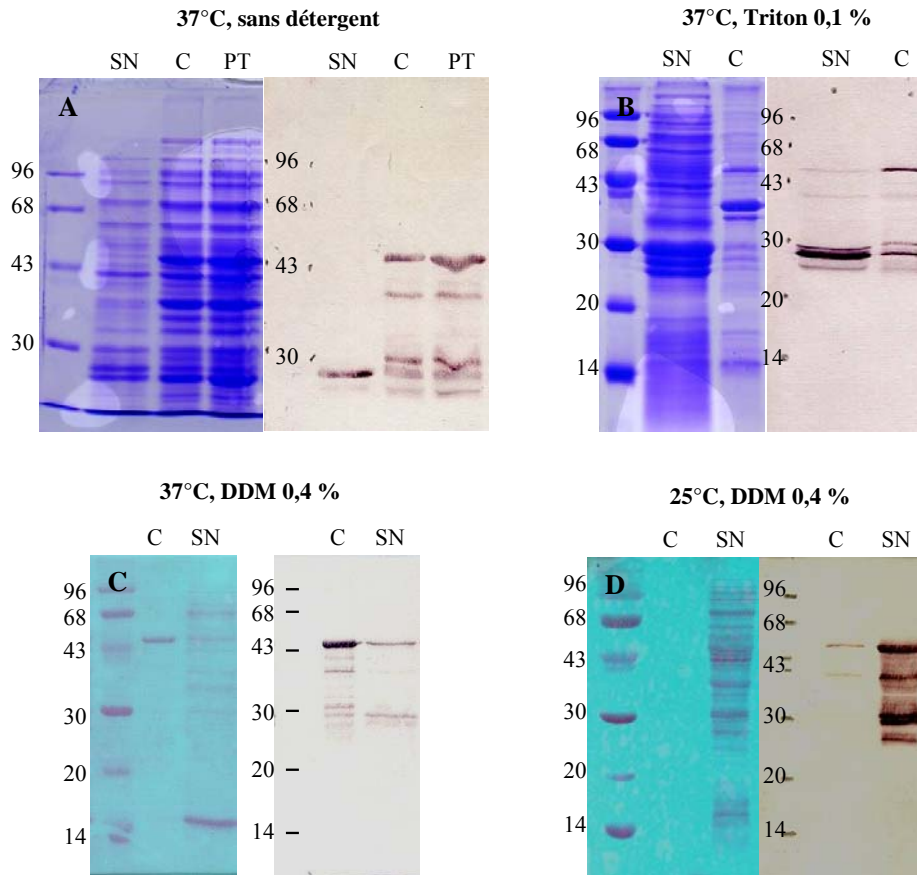
##### 2.1.1. Mise au point de la production et de la solubilisation de la protéine recombinante GST-PtaGLIM1a

La séquence du transcrit *PtaGLIM1a* contient au total 15 codons rares chez *E. coli* (2 CTA, 5 GGA, 3 AGG, 4 AGA, et 1 ATA). Les souches de bactérie Rosetta et Rosetta pLysS synthétisant l'ARNt correspondant à ces codons ont donc été choisies pour la production de la protéine PtaGLIM1a. Succinctement, après transformation, les bactéries sont cultivées à 37°C. Puis la synthèse de la protéine recombinante est induite avec 0,1 mM d'IPTG pendant 2 h à 37°C. Le culot de bactéries est repris dans un tampon de lyse sans détergent. Après centrifugation, les protéines solubles du surnageant et les protéines insolubles du culot sont analysées par migration SDS-PAGE. Les protéines sont ensuite, soit colorées au bleu de Coomassie, soit révélées en Western-blot avec un anticorps monoclonal anti-GST. Tout d'abord, dans la souche Rosetta, nous pouvons observer que la protéine GST-PtaGLIM1a est dégradée après solubilisation des protéines dans le tampon de lyse sans détergent (Fig. 57A). En effet, en Western-blot des polypeptides migrant vers 50, 38, 30, 29 et 27 kDa sont détectés par l'anticorps anti-GST. Les protéines totales de la bactérie directement solubilisées dans du Laemmli, montrent que la protéine est déjà dégradée dans la bactérie Rosetta. Cette dégradation se produit donc lors de la production de la protéine recombinante par la bactérie. La GST-PtaGLIM1a migre vers 50 kDa mais est sous forme insoluble dans le culot. Les produits de dégradation de la protéine GST-PtaGLIM1a sont également insolubles et seule la GST libre migrant vers 29 kDa est soluble (Fig. 57A). Bien qu'insoluble et dégradée, la protéine GST-PtaGLIM1a est cependant produite en grande quantité dans les bactéries.

Dans un second temps, pour résoudre ces problèmes de dégradation et de précipitation de la protéine recombinante GST-PtaGLIM1a, nous avons testé à la fois différentes conditions

### Partie 3. La protéine PtaGLIM1a

de culture de bactéries et différentes méthodes de solubilisation des protéines. Afin d'améliorer la solubilité de la protéine lors de l'extraction de protéines, des détergents comme du Triton X100 à 0,1 % et du dodécylmaltoside (DDM) à 0,4 % ont d'abord été testés.



**Figure 57.** Production et solubilité de la protéine recombinante GST-PtaGLIM1a

La production de la protéine GST-PtaGLIM1a dans la bactérie Rosetta est induite avec 0,1 mM d'IPTG à 25 ou 37°C. Le culot de bactéries est soit directement solubilisé dans du Leammli pour constituer les protéines totales (PT), soit repris dans du tampon de lyse avec ou sans détergent et lysé par 4 cycles de gel/dégel. Le lysat contenant les protéines totales est centrifugé à 16000 g pendant 15 min puis les protéines insolubles du culot (C) sont séparées des protéines solubles du surnageant (SN), et du Leammli est ajouté. Les protéines séparées sur gel SDS-PAGE sont révélées soit par coloration au bleu de coomassie (figures de gauche) soit par Western-blot avec un anticorps monoclonal anti-GST (figures de droite). Les marqueurs de taille sont indiqués en kDa. Les conditions de production et de solubilisation suivantes ont été testées : **(A)** Induction 2 h à 37°C et reprise dans le tampon de lyse sans détergent. **(B)** Induction 2 h à 37°C et reprise dans le tampon de lyse contenant du triton 0,1 %. **(C)** Induction 2 h à 37°C et reprise dans le tampon de lyse contenant du DDM 0,4 %. **(D)** Induction la nuit à 25°C et reprise dans le tampon de lyse contenant du DDM 0,4 %.

De plus, des inhibiteurs de protéases (IP) bactériennes dilués au 1/100<sup>e</sup> et du PMSF à 1 mM ont été ajoutés au tampon de lyse pour éviter la dégradation de la protéine GST-PtaGLIM1a. Ensuite, d'autres conditions de culture des bactéries Rosetta et Rosetta 2 pLysS ont également été testées pour obtenir une protéine plus soluble et réduire sa dégradation. Des températures de cultures plus faibles de 19°C et 25°C ont été appliquées pour ralentir la croissance bactérienne et ainsi diminuer la production de la protéine recombinante insoluble.

### Partie 3. La protéine PtaGLIM1a

De même, l'ajout d'éthanol à 2% dans le milieu de culture provoque un léger stress pour les bactéries et permet de ralentir leurs multiplications. L'utilisation de la bactérie Rosetta 2 pLysS permet de diminuer le niveau basal d'expression du gène GST-PtaGLIM1a dans les bactéries avant induction. Ainsi, en diminuant la vitesse de production de protéines recombinantes, les bactéries tolèrent peut-être mieux la présence d'une protéine exogène et nous espérons de cette manière que la protéine GST-PtaGLIM1a soit moins dégradée et mieux solubilisée. Enfin, de par la présence de deux doigts de zinc au niveau des domaines LIM, la protéine PtaGLIM1a native contient en tout quatre atomes de zinc dans sa structure. Nous avons donc expérimenté l'ajout de zinc sous la forme de ZnSO<sub>4</sub> à 50 µM dans le milieu de culture, pour que les domaines LIM puissent adopter une bonne structure secondaire et un bon repliement tertiaire, favorisant ainsi la solubilité de la protéine recombinante.

L'utilisation de Triton X100 0,1 % (sans IP et PMSF) améliore la solubilité des protéines bactériennes. Nous pouvons en effet observer une plus grande quantité de protéines bactériennes dans le surnageant que dans le culot (Fig. 57B gauche). La solubilité de la protéine recombinante GST-PtaGLIM1a est aussi meilleure mais de façon insuffisante et celle-ci reste fortement dégradée. Une grande quantité de GST libre est visible dans la fraction de protéines solubles (Fig. 57B droite). Le DDM à 0,4 % favorise encore mieux la solubilité des protéines bactériennes ainsi que celle de la protéine GST-PtaGLIM1a. En effet, très peu de protéines bactériennes sont visibles dans le culot de protéines insolubles et elles sont presque toutes solubilisées par le DDM (Fig. 57C gauche). Une plus grande quantité de protéine GST-PtaGLIM1a migrant vers 50 kDa est solubilisée de cette manière. De plus, les IP et le PMSF diminuent la dégradation de la protéine recombinante GST-PtaGLIM1a lors de l'extraction (Fig. 57C droite). Toutefois, les formes dégradées de la protéine GST-PtaGLIM1a sont toujours présentes dans les fractions solubles et insolubles, et sont aussi visibles lorsque l'on solubilise directement le culot de bactéries dans du Laemmli (Fig. 57A et 57C).

A 25°C, les résultats nous montrent qu'après extraction des protéines en présence de DDM à 0,4 %, presque toutes les protéines bactériennes sont solubilisées (Fig. 57D). Mais surtout, nous pouvons constater que la protéine recombinante GST-PtaGLIM1a est nettement plus soluble à 25°C qu'à 37°C. Toutefois, il existe toujours une forte dégradation de la protéine GST-PtaGLIM1a produisant des polypeptides de différentes tailles quelques soient les conditions utilisées. Par ailleurs, il n'existe pas de différences entre les conditions de culture à 19°C et 25°C (Tableau 15). De plus, aucun effet notable du zinc ou de l'éthanol n'a été constaté sur la solubilité ou la dégradation de la protéine GST-PtaGLIM1a. Enfin, la quantité de protéine recombinante produite par la bactérie Rosetta est en général légèrement

### Partie 3. La protéine PtaGLIM1a

supérieure à la quantité produite par Rosetta 2 pLysS (Tableau 15), et j'ai donc choisi Rosetta pour la production de la GST-PtaGLIM1a. En conclusion, ces expériences m'ont permis de choisir les bons paramètres de culture de bactéries et de solubilisation de la protéine recombinante. La protéine GST-PtaGLIM1a est en majorité soluble quand on cultive les bactéries Rosetta à 25°C (induction avec l'IPTG puis culture durant la nuit), et quand on solubilise les protéines avec du DDM 0,4% en présence de PMSF à 1 mM et de IP (Fig. 57D). Bien que la protéine GST-PtaGLIM1a reste dégradée, la forme non dégradée est soluble et produite en quantité suffisante pour la suite des expériences.

**Tableau 15.** Récapitulatif de l'effet des différentes conditions de culture de bactéries testées sur la production, la solubilité et la dégradation de la protéine GST-PtaGLIM1a

Facteurs	Production	Solubilité	Dégradation
37°C	++	+	oui
25 °C	++(+)	+++	oui
19°C	++(+)	+++	oui
éthanol 2 %	pas d'effet	pas d'effet	oui
ZnSO <sub>4</sub> 50 µM	pas d'effet	pas d'effet	oui
Rosetta	+++	pas d'effet	oui
Rosetta pLysS	++	pas d'effet	oui

Les souches de bactéries Rosetta ou Rosetta pLysS sont cultivées dans du milieu LB avec ou sans ajout d'éthanol et de zinc, et à différentes températures : 37, 25 ou 19°C. Les protéines ont été extraites dans du tampon de lyse des bactéries en présence de DDM 0,4 %, de PMSF 1 mM et des inhibiteurs de protéase dilués au 1/100<sup>e</sup>.

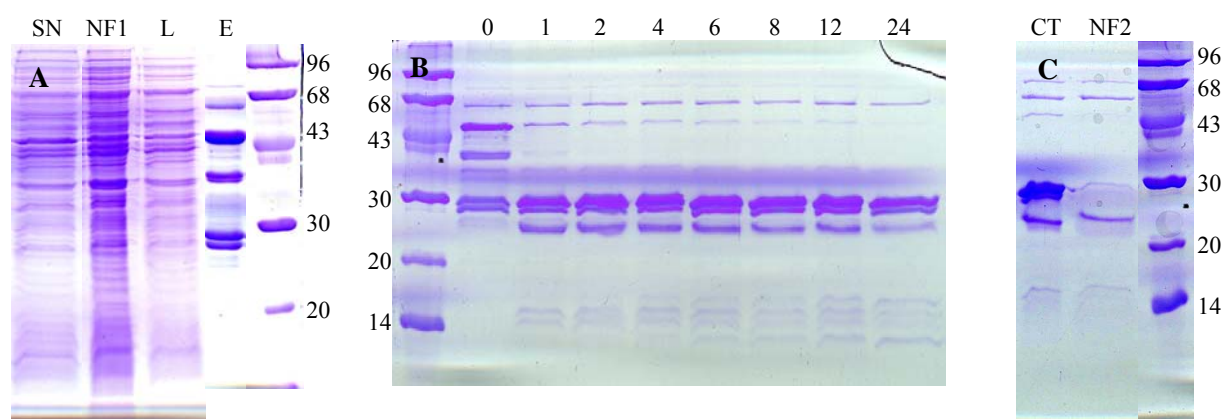
#### 2.1.2. Purification de la protéine recombinante GST-PtaGLIM1a par affinité sur résine de Glutathione Sépharose 4B et clivage de la GST par la thrombine

Après production et solubilisation de la protéine recombinante GST-PtaGLIM1a selon les conditions établies ci-dessus, celle-ci a ensuite été purifiée par chromatographie d'affinité sur une résine de Glutathione Sepharose 4B (Fig. 58A). La colonne a été lavée avec du tampon contenant des sels (NaCl 150 ou 300 mM) et du détergent (Triton 0,1 %) pour éliminer les protéines bactériennes fixées de manière non spécifique. La protéine GST-PtaGLIM1a est éluée avec un excès de glutathion réduit (10 mM). Après élution, nous détectons la protéine recombinante GST-PtaGLIM1a entière vers 50 kDa et des formes dégradée de cette protéine vers 38, et 29 kDa. La GST a été clivée et se situe vers 27 kDa. Mais en plus de la GST-PtaGLIM1a et de ces différentes formes dégradées, deux autres protéines contaminantes de haut poids moléculaire (PM) de 58 et 70 kDa sont également éluées (Fig. 58A).



### Partie 3. La protéine PtaGLIM1a

Le clivage par la thrombine de la protéine recombinante GST-PtaGLIM1a a été testé sur une cinétique de 0 à 24 h à température ambiante ou à 4°C. A température ambiante, le clivage par la thrombine est trop rapide et dès 1 h la protéine PtaGLIM1a commence à être dégradée (résultats non montrés). Sur une cinétique à 4°C, on peut observer que le clivage par la thrombine à 0,1 unité/ml est optimal entre 1 et 2 h ; ceci malgré l'apparition de deux bandes vers 14 kDa qui peuvent correspondre au clivage des deux polypeptides dégradés GST-PtaGLIM1a situés vers 38 kDa (Fig. 58B). La protéine PtaGLIM1a libre est observée vers 25 kDa et reste soluble après clivage de la GST. En Western-blot, la quantité de GST vers 29 kDa augmente en parallèle entre 0 et 2 h de cinétique (résultats non montrés). Après 2 h, la quantité de protéine PtaGLIM1a libre diminue et deux bandes supplémentaires apparaissent aux alentours de 14 kDa suggérant une dégradation de la protéine PtaGLIM1a par un clivage non spécifique par la thrombine. J'ai donc décidé d'utiliser pour le clivage par la thrombine un temps d'incubation de 1h30 à 4°C.



**Figure 58.** Purification de la protéine recombinante GST-PtaGLIM1a par affinité, clivage de la GST par la thrombine et élimination de la GST

(A) Purification de la protéine GST-PtaGLIM1a par affinité sur un gel de Glutathione Sepharose 4B.

Les protéines GST-PtaGLIM1a solubilisées dans le tampon de lyse contenant 0,4 % DDM (SN) sont incubées pendant 1 h avec un gel de Glutathione Sepharose 4B. Le reste des protéines bactériennes ne se fixe pas au gel de Glutathione (NF1). La colonne est ensuite lavée pour éliminer les protéines fixées non spécifiquement (L). La protéine GST-PtaGLIM1a est éluée avec du glutathion à 10 mM (E). Le marqueur de taille est indiqué en kDa.

(B) Cinétique de clivage en heure de la protéine GST-PtaGLIM1a par la thrombine à 0,1 unité/ml et à 4°C.

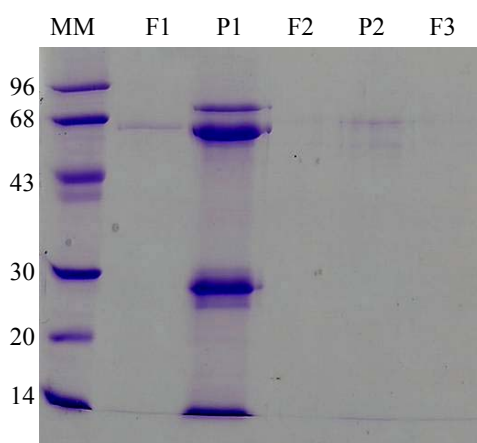
(C) Clivage et élimination de la GST sur gel de Glutathione Sepharose 4B.

La protéine GST-PtaGLIM1a est clivée par la thrombine pendant 1h30 à 4°C (CT) puis la GST libre est fixée aux billes de Glutathione Sepharose 4B pendant 30 min, et la protéine PtaGLIM1a libre se retrouve dans la fraction non fixée au gel de Glutathione (NF2).

Après clivage de la protéine GST-PtaGLIM1a par la thrombine, la GST libre est ensuite éliminée par fixation sur la résine de Glutathione Sepharose 4B. Dans notre cas, malgré une forte diminution du signal GST vers 29 kDa dans l'échantillon non fixé (NF2), la totalité des protéines GST ne semble pas s'être fixée à la résine. Par ailleurs, la quantité de protéine PtaGLIM1a soluble diminue après incubation avec la résine suggérant une

### Partie 3. La protéine PtaGLIM1a

dégradation ou une précipitation de la protéine lors de cette étape (Fig. 58C). De plus les deux protéines contaminantes de haut poids moléculaire vers 72 kDa et 58 kDa sont en quantité non négligeable par rapport à la quantité de protéine PtaGLIM1a libre. Ces deux bandes de protéines contaminantes déjà observées après élution de la protéine fusion GST-PtaGLIM1a, sont peut-être des dimères de la protéine PtaGLIM1a formés par des ponts disulfures, ou bien correspondraient aux protéines chaperonnes d'*E. coli* dnaK (70 KDa) et GroEL (58 KDa). En effet, il a déjà été observé pour d'autres protéines fusionnées à la GST une co-purification des protéines dnaK et GroEL impliquées dans le bon repliement des protéines en cours de synthèse dans la bactérie (Buchberger *et al.*, 1996 ; Thain *et al.*, 1996 ; Rohman & Harrison-Lavoie, 2000).



**Figure 59.** Analyse des fractions éluées après tamisage moléculaire de la protéine GST-PtaGLIM1a sur colonne de gel filtration « Superdex 75 »

Les fractions ont été rassemblées et concentrées par précipitation au TCA avant chargement sur gel SDS-PAGE et coloration au bleu de Coomassie. P1 et P2 : fractions respectives des premier et deuxième pics de concentration en protéine (Abs 280 nm). F1, F2 et F3 : fractions respectivement recueillies avant le premier pic, entre les deux pics, et après le deuxième pic.

#### 2.1.3. Test de l'élimination des protéines contaminantes co-purifiées avec la protéine GST-PtaGLIM1a

Nous avons ensuite cherché à éliminer ces deux protéines contaminantes par tamisage moléculaire sur une colonne de gel filtration « Superdex 75 » mais sans succès. Lors de la mesure de la concentration en protéines (Abs 280 nm) des fractions à la sortie de la colonne de gel filtration, deux pics sur la courbe de mesure apparaissent : un premier pic semblant contenir deux pics très proches et un deuxième pic d'amplitude très faible (résultats non montrés). Or l'analyse en gel SDS-PAGE des fractions recueillies nous indique que le premier pic contient les protéines contaminantes de haut poids moléculaire vers 58 kDa et 72 kDa, la

### Partie 3. La protéine PtaGLIM1a

GST vers 29 kDa et un polypeptide vers 14 kDa (Fig. 59). De plus, le deuxième pic ne contient aucune protéine, et la protéine PtaGLIM1a semble avoir disparu lors du tamisage moléculaire. Bien que ces résultats soient difficilement interprétables, ceci suggère une éventuelle précipitation de la protéine PtaGLIM1a dans la colonne de gel filtration. Ces deux protéines contaminantes ont été identifiées en spectrométrie de masse MS/MS. Ce sont deux isoformes de GroEL, une protéine chaperonne de la bactérie. Par sa fixation aux polypeptides en cours de synthèse, elle aide à la bonne solubilisation et conformation des protéines bactériennes notamment en cas de stress thermique (Ellis, 2005).

Une autre manière d'éliminer ces deux protéines est de laver la colonne Glutathione Sepharose 4B avec de l'ATP avant l'élution par le glutathion (Thain *et al.*, 1996 ; Rohman & Harrison-Lavoie, 2000). Dans certains cas, l'ATP permet à GroEL de libérer le polypeptide capturé. Le lavage de la colonne avec de l'ATP (5 et 10 mM) est efficace et permet bien l'élimination de GroEL (résultats non montrés). Cependant, la protéine GST-PtaGLIM1a sans GroEL est instable. En effet peu de protéine GST-PtaGLIM1a est éluee après ces lavages laissant penser qu'elle précipite dans la colonne (résultats non montrés). De plus, lors de la dialyse et du clivage à la thrombine, une grande quantité de protéine GST-PtaGLIM1a précipite, et au final la protéine PtaGLIM1a libre et soluble disparaît après clivage à la thrombine. Dans ces conditions GroEL semble indispensable à la bonne conformation de PtaGLIM1a et sans la GST favorisant partiellement la solubilité de PtaGLIM1a, cette dernière est totalement insoluble. Enfin, en présence de GroEL et sans détergent, les protéines GST-PtaGLIM1a et surtout PtaGLIM1a ont aussi une forte tendance à précipiter. J'ai donc choisi une autre stratégie pour la production de protéine recombinante PtaGLIM1a.

#### 2.2. Production de protéines recombinante 6His-PtaGLIM1a :

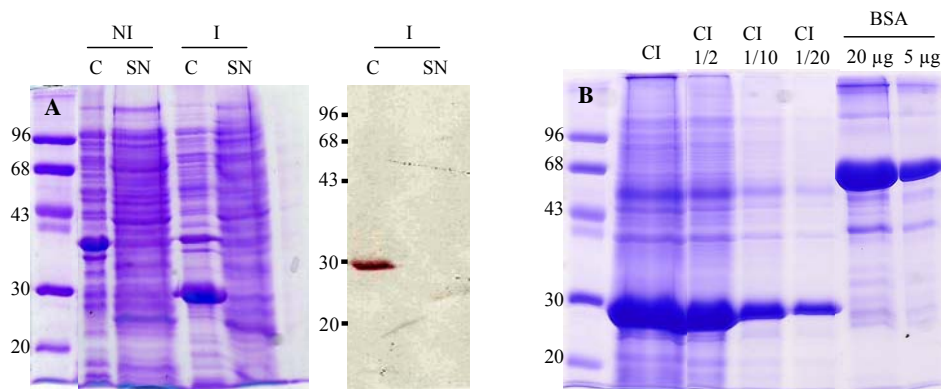
##### 2.2.1. Mise au point de la production et de la solubilisation de protéines recombinantes 6His-PtaGLIM1a dans différentes souches d'*E. coli*

La protéine recombinante GST-PtaGLIM1a donne une protéine PtaGLIM1a insoluble difficile à purifier. J'ai choisi donc d'utiliser l'étiquette 6His pour la production et la purification de la protéine PtaGLIM1a. En effet, elle offre un avantage : il n'est pas nécessaire de cliver l'étiquette 6His pour la production d'anticorps et l'étude fonctionnelle de la protéine. Différentes souches bactériennes ont été testées : BL21, BL21 pLysS, C41, C43, Origami,

### Partie 3. La protéine PtaGLIM1a

Origami pLysS, Rosetta et Rosetta pLysS et Tuner pLysS. Après transformation, les bactéries ont été cultivées à 24°C ou à 37°C. La production de protéine recombinante est induite avec 0,1 mM d'IPTG pendant 4h à 37°C ou durant la nuit à 24°C. Le culot a été repris dans un tampon de lyse sans détergent. Après centrifugation pendant 15 min à 16000 g, le culot de protéines insolubles et le surnageant de protéines solubles sont analysés par migration SDS-PAGE. Les protéines sont ensuite soit colorées au bleu de Coomassie soit révélées en Western Blot avec un anticorps monoclonal anti-6His.

Après vérification en Western-blot, aucune souche de bactérie ne produit de protéine 6His-PtaGLIM1a sous forme soluble quelque soit la température de culture. Les souches de bactéries Rosetta et C43 ne produisent pas la protéine recombinante (résultats non montrés). Pour les autres souches bactériennes, la protéine recombinante 6His-PtaGLIM1a est produite en grande quantité, mais est insoluble dans le culot. J'ai donc décidé de produire la protéine 6His-PtaGLIM1a sous la forme de corps d'inclusion (CI) et de les purifier ensuite par des lavages. Pour cela la souche BL21 pLysS a été choisie. Elle a une croissance rapide à 37°C et peut être facilement lysée. L'induction de la production de la protéine PtaGLIM1a est bien visible après coloration au bleu de Coomassie avec un fort signal à 27 kDa dans le culot de protéine insoluble (Fig. 60A). L'identité de cette protéine a été vérifiée par Western-blot avec un anticorps anti-6His.



**Figure 60.** Production de la protéine 6His-PtaGLIM1a sous forme de corps d'inclusion dans la bactérie BL21 pLys S et purification des corps d'inclusion

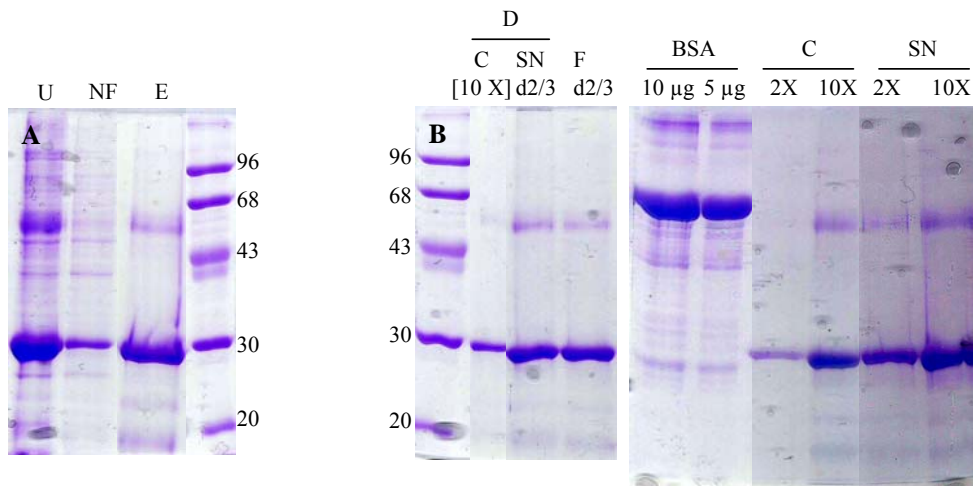
(A) La production de la protéine 6His-PtaGLIM1a dans la bactérie BL21 pLysS est non induite (NI) ou induite (I) par de l'IPTG 0,1 mM pendant 4 h à 37°C. Le culot de bactéries est repris dans du tampon de lyse sans détergent. Après 15 min de centrifugation à 16000 g, les protéines insolubles du culot (C) sont séparées des protéines solubles du surnageant (SN). Les protéines sont séparées par SDS-PAGE puis sont révélées par coloration au bleu de Coomassie ou par Western-blot avec un anticorps monoclonal anti-6His. Le marqueur de taille en indiqué en kDa.

(B) Les corps d'inclusion sont lavés quatre fois avec du tampon de lyse contenant du triton 0,5 %. 20 µl du culot final de corps d'inclusion (CI) non dilué ou dilué au 1/2e, 1/10e et 1/20e est révélé au bleu de Coomassie. 20 µg et 5 µg d'albumine de sérum bovin (BSA) sont déposés sur gel pour estimer la quantité de protéines 6His-PtaGLIM1a dans les corps d'inclusion.

### Partie 3. La protéine PtaGLIM1a

#### 2.2.2. Purification de la protéine recombinante 6His-PtaGLIM1a sous forme de corps d'inclusion puis par affinité sur une résine de cobalt

Pour la production de la protéine recombinante 6His-PtaGLIM1a en grande quantité sous forme de corps d'inclusion, deux litres de culture ont été inoculés. Afin de purifier la protéine 6His-PtaGLIM1a, les corps d'inclusion ont été lavés avec du tampon de lyse contenant du Triton X100 0,5 % pour éliminer les protéines membranaires bactériennes. Au final, par comparaison des différentes dilutions des corps d'inclusion avec la gamme d'albumine de sérum bovin (BSA), la purification de la protéine 6His-PtaGLIM1a est estimée à au moins 90 % avec une concentration de 2,5 mg/ml (Fig. 60B). Ce degré de pureté est suffisant pour lancer la production d'anticorps dirigés spécifiquement contre la protéine recombinante 6His-PtaGLIM1a.



**Figure 61.** Purification par affinité après solubilisation dans l'urée de la protéine 6His-PtaGLIM1a, élimination des agents dénaturants par dialyse et tests de solubilité de la protéine 6His-PtaGLIM1a

**(A)** Purification par affinité de la protéine 6His-PtaGLIM1a sur une résine de Cobalt

Les corps d'inclusions sont solubilisés dans du tampon SL contenant de l'urée 8 M (U), puis la protéine 6His-PtaGLIM1a est fixée à une résine de Cobalt BD TALON™. Les protéines bactériennes non fixées (NF) sont éliminées. La résine est lavée et les protéines 6His-PtaGLIM1a sont éluées avec du tampon SL contenant 400 mM d'imidazole et 0,8 M d'arginine (E). Le marqueur de taille est indiqué en kDa.

**(B)** Test de la solubilité de la protéine PtaGLIM1a après dialyse

Les protéines 6His-PtaGLIM1a sont dialysées trois fois dans du tampon de dialyse contenant du glycérol 5 %, du ZnSO<sub>4</sub> 1 mM, et des concentrations décroissantes d'arginine (0,4 M, 0,2 M et 0 M). La bonne conformation des protéines 6His-PtaGLIM1a du dialysat final (D) est testée par filtration à 0,22 µm (F) et en concentrant 500 µl de dialysat 2 fois (2X) pour un volume final de 290 µl, puis 10 fois (10X) pour un volume final de 65 µl. La solubilité des protéines est vérifiée par centrifugation à 16000 g pendant 15 min. Le culot (C) contient les protéines insolubles (concentrées 10 fois pour le dialysat) et le surnageant (SN) contient les protéines solubles. (Le surnageant du dialysat et le filtrat sont dilués au 2/3). 10 µl de protéines est chargé sur gel SDS-PAGE.

Cependant, pour une meilleure pureté et surtout pour obtenir une protéine soluble fonctionnelle, j'ai entrepris de solubiliser la protéine 6His-PtaGLIM1a par dénaturation dans l'urée, de la purifier par chromatographie d'affinité sur une résine de cobalt, et si possible, de la renaturer dans une bonne conformation structurale. La protéine 6His-PtaGLIM1a contenue

### Partie 3. La protéine PtaGLIM1a

dans les corps d'inclusion a été solubilisée dans du tampon SL contenant de l'urée 8 M, pour être ensuite purifiée par fixation sur résine de Cobalt BD TALON™. La majorité des protéines bactériennes restantes a été éliminée (Fig. 61A) et après élution avec 400 mM d'imidazole, le degré de pureté de PtaGLIM1a est plus élevé (environ 95 %). En plus de la protéine 6His-PtaGLIM1a migrant vers 29 kDa, nous pouvons également observer dans l'éluat un polypeptide de taille élevée vers 50 kDa et deux autres polypeptides de taille plus faible aux alentours de 24 et 14 kDa. Comme nous le verrons plus tard, le premier polypeptide pourrait correspondre à un dimère de la protéine 6His-PtaGLIM1a et les seconds seraient le résultat d'une dégradation de la protéine recombinante.

#### 2.2.3. Tests de renaturation et de solubilisation de la protéine recombinante 6His-PtaGLIM1a

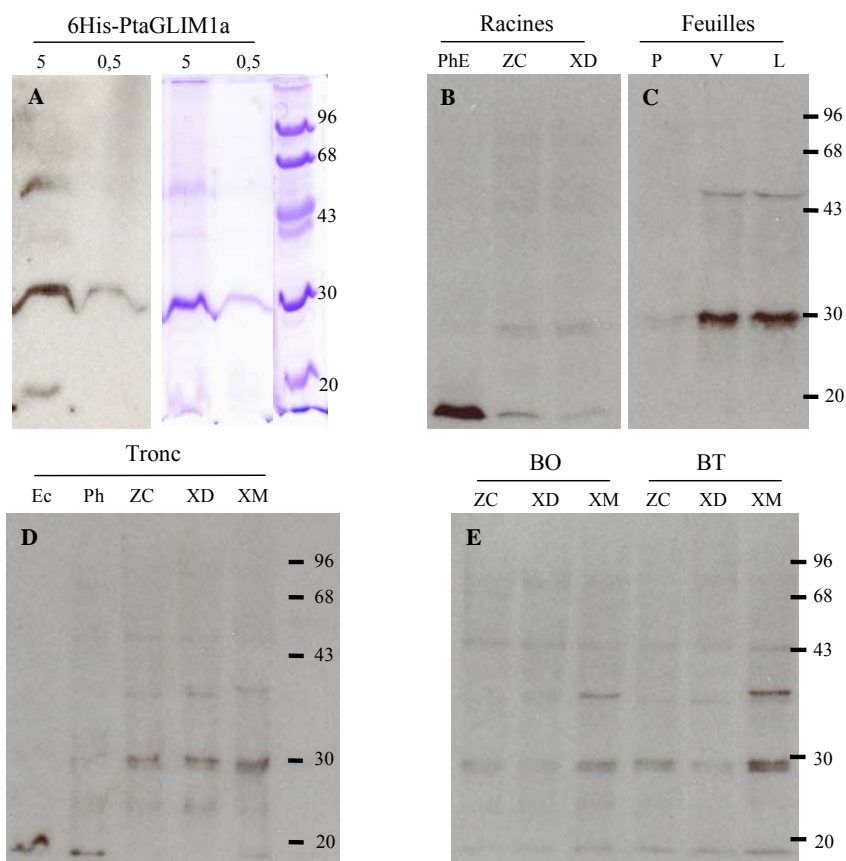
La renaturation de la protéine débute à partir de l'élution sur colonne en remplaçant l'urée par un agent dénaturant moins fort, l'arginine à une concentration de 0,8 M. La renaturation s'est poursuivie par des dialyses successives dans un tampon de dialyse ayant des concentrations décroissantes en arginine : 0,4 M, 0,2 M et 0 M. Du glycérol 5 % a été ajouté dans les tampons de dialyse pour protéger la protéine. Egalement, un excès de zinc (1 mM de ZnSO<sub>4</sub>) a été ajouté pour la renaturation des quatre doigts de zinc de la protéine. La dernière dialyse (sans arginine) est réalisée dans un tampon contenant 10 % de glycérol pour la conservation de la protéine. La solubilité de la protéine 6His-PtaGLIM1a a été testée par filtration à 0,22 µm. Puis, j'ai analysé le caractère soluble ou insoluble des protéines lorsqu'elles sont concentrées. La concentration des protéines a été effectuée sur une unité à filtration de 10 kDa, à partir d'un volume initial de 500 µl de dialysat. Après centrifugation 15 min à 16000 g, on peut voir sur gel SDS-PAGE et coloration au bleu de Coomassie, les protéines précipitées dans le culot et les protéines solubles dans le surnageant. La concentration en protéines des surnageants a été estimée par comparaison avec une gamme de BSA et par des mesures d'absorbance à 280 nm. Après les dialyses, les protéines 6His-PtaGLIM1a sont majoritairement solubles (Fig. 61B) avec une concentration de 0,25 mg/ml. La quantité de protéines solubles après filtration à 0,22 µm ne semble pas être affectée, car celles-ci ont toujours une concentration de 0,25 mg/ml. Lors de la concentration des protéines, celles-ci ont tendance à précipiter dans un premier temps : Quand le volume diminue de moitié, la concentration en protéine est identique à celle de départ voire augmente légèrement (0,31 mg/ml). Ensuite la teneur en protéine augmente quand on poursuit la concentration en

diminuant le volume 10 fois (0,77 mg/ml). Ceci indique qu'au moins la moitié des protéines n'ont pas un bon repliement et précipitent quand elles se trouvent concentrées. Les conditions de renaturation de la protéine 6His-PtaGLIM1a sont donc à améliorer pour atteindre une concentration en protéine soluble plus élevée de l'ordre de 2 mg/ml. L'obtention de cette protéine avec un bon repliement structural sera nécessaire pour la réalisation d'expériences de gel retard visant à montrer que la protéine se fixe sur l'ADN. D'autre part, nous avons envoyé la protéine recombinante 6His-PtaGLIM1a purifiée par affinité à la société AGRO-BIO pour la production d'anticorps anti-6His-GLIM1a par deux lapins.

#### 2.3. Détection avec les anticorps A16 des protéines LIM de peuplier appartenant au groupe LIM1

Dans un premier temps, nous avons utilisé les anticorps A16 (lapin 1) afin d'étudier l'expression des protéines à domaine LIM chez le peuplier. Ces anticorps A16 sont dirigés contre le peptide P16 « CDQLFKKTGSLDKSFE » situé dans la région inter-LIM de la protéine HaWLIM1, et a été purifié par affinité contre ce peptide (Brière *et al.*, 2003). Cette séquence peptidique est quasiment identique aux régions inter-LIM des protéines de peuplier du groupe LIM1 : PtaGLIM1a et b, PtaWLIM1a et b, et Pta $\beta$ LIM1a et b (Fig. 56A). Les anticorps correspondants peuvent théoriquement reconnaître ces six protéines chez le peuplier et nous donner un aperçu de l'abondance globale des protéines LIM1 chez le peuplier. J'ai tout d'abord testé la reconnaissance de ces anticorps A16 vis à vis de la protéine recombinante 6His-PtaGLIM1a contenue dans les corps d'inclusion. Nous pouvons constater que ces anticorps détectent bien des quantités de 5 et 0,5  $\mu$ g de protéine 6His-PtaGLIM1a (Fig. 62A). De plus trois polypeptides sont révélés : un correspondant à la taille de la protéine 6His-PtaGLIM1a vers 29 kDa, et deux autres polypeptides de 50 et 14 kDa. Celui vers 50 kDa pourrait correspondre à la formation de dimères de protéines 6His-PtaGLIM1a, et l'autre vers 14 kDa semble plutôt provenir d'une dégradation de la protéine.

### Partie 3. La protéine PtaGLIM1a



**Figure 62.** Immunodétection avec les anticorps A16 de la protéine recombinante 6His-PtaGLIM1a et des protéines du groupe LIM1 dans des extraits de protéines totales issues de différents tissus de peuplier (*Populus tremula* × *P. alba*)

La reconnaissance de la protéine PtaGLIM1a par les anticorps A16 (lapin 1) purifié par affinité a été testée par Western-blot sur 5 à 0,5 µg de protéines recombinantes 6His-PtaGLIM1a sous forme de corps d'inclusion (A). En parallèle 5 à 0,5 µg de protéines 6His-PtaGLIM1a sont colorées au bleu de coomassie après migration SDS-PAGE. L'expression de la protéine PtaGLIM1a a été analysée par Western-blot sur des protéines totales extraites à partir de différents tissus prélevés sur la racine principale (B), les feuilles (C) et sur le tronc d'un peuplier non incliné (D) cultivés en pépinière pendant quatre ans. Sur le tronc d'un peuplier incliné (E) âgé de quatre ans, les échantillons de bois opposé (BO) et de bois tendu (BT) ont été analysés séparément. Au niveau du tronc, le phloème (Ph) a été séparé de l'écorce (Ec), et la zone cambiale (ZC), le xylème en développement (XD) et le xylème mature (XM) ont été prélevés au scalpel. Le jeune xylème (JX), la zone cambiale (ZC) et le phloème incluant l'épiderme (PhE) ont été prélevés sur la racine principale. Au niveau des feuilles, les pétioles (P), les vaisseaux (V) et le limbe (L) ont été séparés au scalpel. 10 µg de protéines totales ont été chargées sur gel SDS-PAGE. Les anticorps A16 purifiés par affinité contre le peptide P16 ont été utilisés au 1/200e, et les membranes révélées par luminescence chimique ont été exposées 30 secondes sous film autoradiographique. La masse moléculaire du marqueur de taille est indiquée en kDa.

Sur des extraits de protéines totales de peuplier, les anticorps A16 reconnaissent plusieurs polypeptides (Fig. 62B à 62E). Deux polypeptides majeurs très proches et migrant vers 29 kDa sont détectés et cette taille ressemble à celles observées des protéines du groupe LIM1 connues (Baltz *et al.*, 1999 ; Mundel *et al.*, 2000 ; Brière *et al.*, 2003). Ce polypeptide vers 29 kDa est très abondant dans les feuilles au niveau du limbe et des vaisseaux (Fig. 62C), et globalement sa quantité augmente dans le bois avec la différenciation du xylème (Fig. 62D). Enfin, il apparaît plus abondant dans le xylème mature de bois tendu que dans celui de



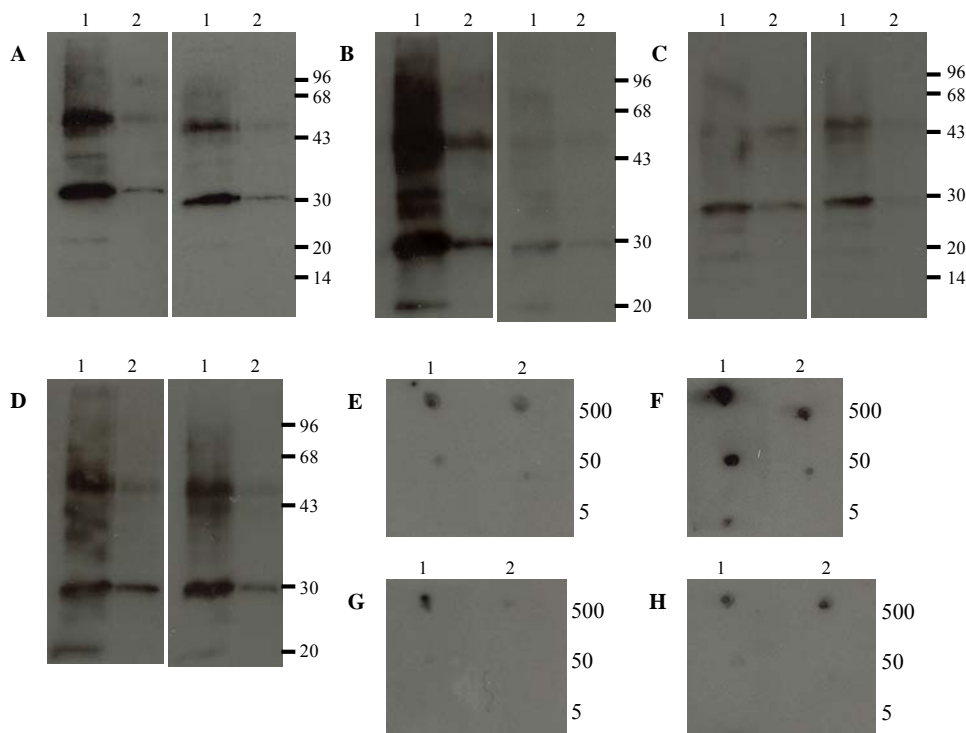
bois opposé (Fig. 62E) et il est difficilement détectable dans le xylème secondaire de la racine (Fig. 62B). Deux autres polypeptides de plus grandes tailles sont observés vers 36 kDa et 45 kDa et pourraient correspondre à des modifications post-traductionnelles de ces protéines LIM1, telles que la formation de dimères ou des phosphorylations et glycosylations. De façon intéressante, le polypeptide vers 45 kDa est plus abondant dans les vaisseaux et limbes de feuilles que dans le bois et inversement le polypeptide vers 36 kDa est plus abondant dans le bois et spécialement dans le xylème mature que dans les feuilles (Fig. 62C et 62E). Enfin, un polypeptide d'une taille inhabituellement basse vers 20 kDa est détecté par les anticorps A16. De manière surprenante, au niveau des racines et du tronc, ce polypeptide de 20 kDa n'est détecté que dans les tissus de l'écorce, du phloème et dans une moindre mesure du cambium (Fig 62B et 62D). Il est très abondant dans le phloème et l'épiderme racinaire. Ainsi, l'utilisation de ces anticorps A16 nous donne des résultats intéressants mais difficilement interprétables car les six protéines LIM1 de peuplier peuvent être reconnues.

#### *2.4. Estimation du titre des anticorps anti-6His-GLIM1a et anti-P-GLIM1a sur la protéine recombinante 6His-PtaGLIM1a et sur le peptide P-GLIM1a*

Nous avons décidé de produire et de purifier la protéine recombinante 6His-PtaGLIM1a afin d'obtenir des anticorps dirigés spécifiquement contre cette protéine. En effet, les anticorps A16 ne sont pas spécifiques de la protéine PtaGLIM1a et peuvent reconnaître les différentes protéines du groupe LIM1. De plus, afin de distinguer les deux protéines PtaGLIM1a et b fortement similaires entre elles, nous avons décidé de produire des anticorps dirigés contre le peptide P-GLIM1a situé dans la région C-terminale de la protéine PtaGLIM1a. Ces anticorps seront utilisés en Western blot pour l'analyse de l'expression de la protéine PtaGLIM1a dans les différentes parties de l'arbre.

La spécificité et le titre des anticorps anti-6His-GLIM1a (lapins 1 et 2) et anti-P-GLIM1a (lapin 1 et 2) reçus par la société AGRO-BIO ont tout d'abord été testés en Western-blot sur la protéine recombinante 6His-PtaGLIM1a purifiée par affinité sur résine de cobalt et en Dot-blot sur le peptide de synthèse P-GLIM1a. Les sérums bruts non purifiés ont été dilués au 1/1000<sup>e</sup> ou au 1/5000<sup>e</sup> avant hybridation, puis révélés par luminescence chimique et exposition des membranes sous films autoradiographiques pendant des temps variables de 5 sec à 1 min.

### Partie 3. La protéine PtaGLIM1a



**Figure 63.** Immunodétection de la protéine recombinante 6His-PtaGLIM1a par Western-blot, et du peptide P-GLIM1a par Dot-blot avec les anticorps anti-P-GLIM1a et anti-6His-GLIM1a (lapin 1 et 2)

La protéine recombinante 6His-PtaGLIM1a purifiée par affinité sur résine de cobalt a été diluée au 1/10e (1) ou au 1/100e (2) avant migration SDS-PAGE (A à D). Les membranes ont été hybridées avec les sérums bruts anti-P-GLIM1a lapin 1 (A), anti-P-GLIM1a lapin 2 (B), anti-6His-GLIM1a lapin 1 (C), et anti-6His-GLIM1a lapin 2 (D) dilués au 1/1000e (panel de gauche) ou au 1/5000e (panel de droite). Après révélation par luminescence chimique, les membranes ont été respectivement exposées pendant 1 min (A), 15 sec (B), et 5 sec (C et D) sous film autoradiographique. Le marqueur de taille est indiqué en kDa.

Le peptide P-GLIM1a a été déposé sur des membranes en Dot-blot pour des quantités finales de 500, 50 et 5 ng. Les membranes ont été hybridées avec les sérums bruts anti-P-GLIM1a lapin 1 (E), anti-P-GLIM1a lapin 2 (F), anti-6His-GLIM1a lapin 1 (G), et anti-6His-GLIM1a lapin 2 (H) dilués au 1/1000e (1) ou au 1/5000e (2). Après révélation par luminescence chimique, les membranes Dot-blot ont été exposées pendant 1 min.

Comme dans le cas des anticorps A16, les quatre sérums d'anticorps reconnaissent bien les différents polypeptides de la protéine 6His-PtaGLIM1a vers 29, 50 et 14 kDa (Fig. 63A à 63D). Ces trois polypeptides correspondraient respectivement à la protéine recombinante PtaGLIM1a, à sa forme modifiée par dimérisation, et à sa forme dégradée. Nous pouvons constater d'après les signaux observés et le temps d'exposition des membranes que les anticorps anti-PtaGLIM1a contenus dans les sérums sont plus ou moins concentrés selon le type d'antigène injecté (protéine 6His-PtaGLIM1a entière ou peptide de synthèse P-GLIM1a) et selon les lapins immunisés. Les sérums bruts issus de lapins immunisés avec la protéine entière sont plus concentrés en anticorps reconnaissant la protéine 6His-PtaGLIM1a (temps d'exposition des membranes de 5 sec) que les sérums bruts issus de lapins immunisés avec le peptide (temps d'exposition de 15 sec et 1 min). De même, quel que soit l'antigène reçu, les lapins 2 ont produit plus d'anticorps dirigés contre la protéine 6His-PtaGLIM1a que

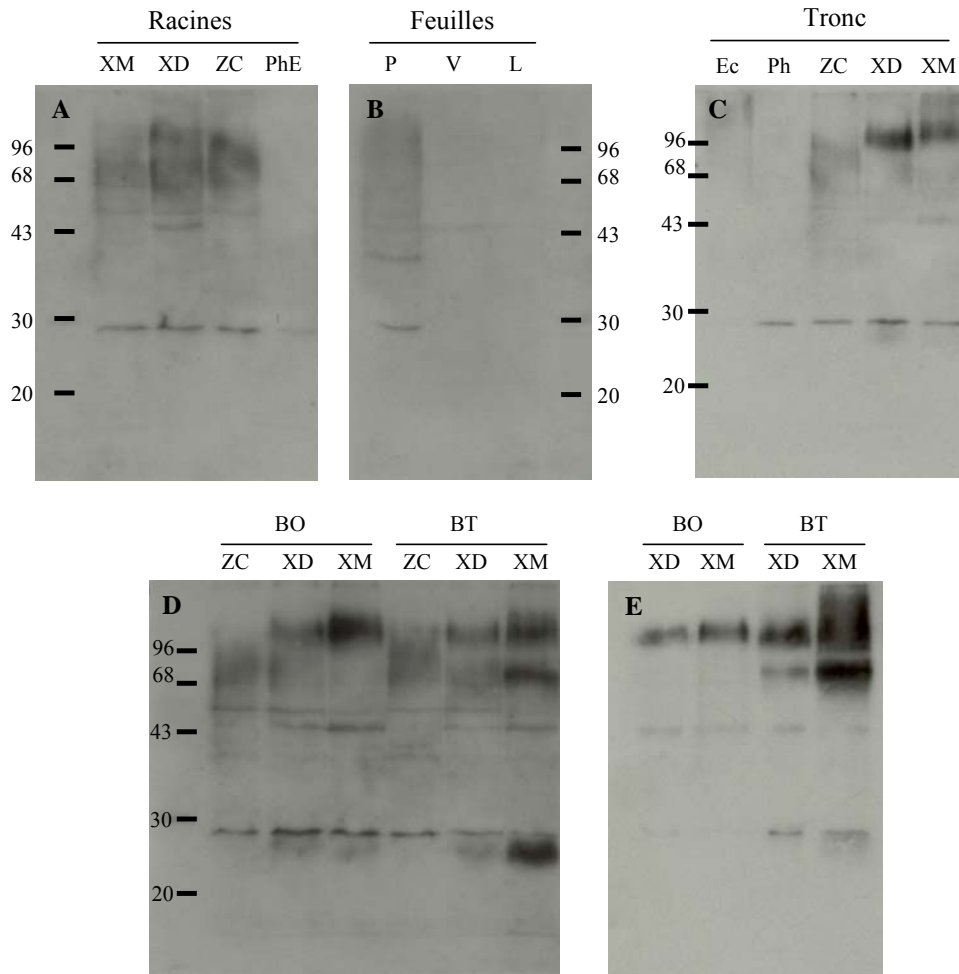
les lapins 1. Ceci est très marqué dans le cas des lapins immunisés avec le peptide de synthèse (Fig. 63A et 63B). Enfin, comme nous pouvons nous y attendre, les analyses en Dot-blot nous indiquent que la reconnaissance du peptide de synthèse P-GLIM1a par les anticorps est globalement plus forte quand on utilise les sérums bruts anti-P-GLIM1 que les sérums bruts anti-6His-GLIM1a (Fig. 63E à 63H). De plus, comme dans le cas du Western-blot, le sérum brut anti-P-GLIM1a du lapin 2 est plus concentré en anticorps spécifiquement dirigés contre le peptide de synthèse que celui du lapin 1 (Fig. 63E et 63F). De même le sérum brut anti-6His-GLIM1a du lapin 2 donne un signal légèrement plus fort en Dot-blot que celui du lapin 1 (Fig. 63G et 63H).

#### *2.5. Immunodétection des protéines PtaGLIM1a de peuplier avec les anticorps anti-P-GLIM1a*

Le sérum brut d'anticorps anti-P-GLIM1a du lapin 2 a été utilisé dans un premier temps pour les analyses en Western-blot car il présente un meilleur titre que celui du lapin 1 et les anticorps anti-P-GLIM1a produits ont plus de chance d'être spécifiques de la protéine PtaGLIM1a que ceux dirigés contre la protéine entière. En effet, les protéines PtaGLIM1a et b présentent 95 % d'identité en acides aminés entre elles et 76 à 78 % d'identité avec les protéines PtaWLIM1a et b (Tableau 10). Ainsi, les anticorps dirigés contre la protéine entière risquent de reconnaître de manière non spécifique la protéine PtaGLIM1b, et de façon moindre les protéines PtaWLIM1a et b.

Sur des extraits de protéines totales de peuplier, le sérum brut d'anticorps anti-P-GLIM1a reconnaît également deux polypeptides vers 29 et 45 kDa (Fig. 64A à 64E). Mais le plus étonnant est la détection de deux à trois polypeptides ayant une migration diffuse et des masses moléculaires très élevées. Un polypeptide migre vers 70 kDa et les deux autres polypeptides migrent vers des masses moléculaires supérieures à 70 et 100 kDa. Le polypeptide migrant vers 29 kDa est détecté dans les tissus du phloème jusqu'au xylème mature prélevés sur le tronc d'un arbre droit et âgé de quatre années, avec une quantité légèrement plus forte dans le xylème en développement (Fig. 64C). De façon surprenante, ce polypeptide de 29 kDa laisse place à un autre polypeptide à migration diffuse et de taille légèrement plus faible vers 27 kDa, uniquement au niveau du xylème mature de bois tendu d'un arbre adulte (Fig. 64D). Le polypeptide de 29 kDa est détecté uniquement dans le bois tendu de jeunes peupliers cultivés en serre (Fig. 64E), et est aussi observé dans la zone cambiale et le xylème de racines d'arbres adultes (Fig. 64A).

### Partie 3. La protéine PtaGLIM1a



**Figure 64.** Immunodétection avec les anticorps anti-P-GLIM1a de la protéine PtaGLIM1a dans des extraits de protéines totales issues de différents tissus de peuplier (*Populus tremula* × *P. alba*)

L'expression de la protéine PtaGLIM1a a été analysée par Western-blot sur des protéines totales extraites à partir de différents tissus prélevés sur la racine principale (A), les feuilles (B), et le tronc d'un peuplier non incliné (C) cultivés en pépinière pendant quatre ans. Sur le tronc d'un peuplier incliné et cultivé en pépinière pendant quatre ans (D) ou sur la tige de trois jeunes peupliers inclinés et cultivés en serre pendant trois mois (E), les échantillons de bois opposé (BO) et de bois tendu (BT) ont été analysés séparément. Au niveau du tronc ou de la tige, le phloème (Ph) a été séparé de l'écorce (Ec), et la zone cambiale (ZC), le xylème en développement (XD) et le xylème mature (XM) ont été prélevés au scalpel. Le cylindre central constitué de xylème mature (XM), le xylème en développement (XD), la zone cambiale (ZC) et le phloème incluant l'épiderme (PhE) ont été prélevés sur la racine principale. Au niveau des feuilles, les pétioles (P), les vaisseaux (V) et le limbe (L) ont été séparés au scalpel. 20 µg de protéines ont été chargées sur gel SDS-PAGE. Le sérum brut anti-P-GLIM1a a été utilisé au 1/1000e, et les membranes révélées par luminescence chimique ont été exposées 10 à 15 secondes sous film autoradiographique. La masse moléculaire du marqueur de taille est indiquée en kDa.

De manière intéressante, la masse moléculaire des polypeptides migrant de façon diffuse et à des tailles élevées (supérieures à 70 et 100 kDa) augmente en fonction de la différenciation du xylème dans le tronc d'arbres adultes inclinés ou droit, et quel que soit le type de bois opposé ou tendu analysé (Fig. 64C et 64D). Par ailleurs, il est remarquable d'observer l'apparition d'un polypeptide à migration diffuse vers 70 kDa uniquement dans le xylème de bois tendu et de façon très marquée au niveau du xylème mature (Fig. 64D et 64E). Ce polypeptide de 70 kDa est très abondant dans le xylème mature de bois de tension chez de

jeunes peupliers cultivés en serre (Fig. 64E). Ces résultats nous montrent que l'expression de la protéine PtaGLIM1a est également induite dans le xylème de bois tendu. Dans les racines, nous observons aussi des polypeptides de taille très élevée de 70 kDa jusqu'à plus de 100 kDa, mais leurs migrations sont trop étalées pour être interprétables (Fig. 64A). Au niveau des feuilles, le sérum brut anti-P-GLIM1a détecte faiblement la protéine PtaGLIM1a uniquement au niveau des pétioles (Fig. 64B). Enfin, les anticorps anti-P-GLIM1a ne détectent aucune protéine PtaGLIM1a dans les organes reproducteurs mâle (étamines et pollen) et femelle (ovaires, pistils et stigmates) de peuplier, ainsi que dans les graines et les fibres de coton (résultats non montrés).

## 3 Discussion

### 3.1. Les protéines PtaGLIM1a pourraient former des dimères de protéine *in vitro* et chez le peuplier

Le profil de migration en gel SDS-PAGE de la protéine 6His-PtaGLIM1a purifiée par affinité sur résine de cobalt est surprenant. En effet, après coloration au bleu de Coomassie ou détection avec les anticorps A16, anti-6His-GLIM1a et anti-P-GLIM1a, trois polypeptides de différente taille (14, 29 et 50 kDa) sont révélés (Fig. 62 et 63). Le polypeptide détecté vers 14 kDa n'est pas produit par la bactérie (Fig. 60), et pourrait correspondre à une dégradation de la protéine par clivage lors des étapes de purification. Trois polypeptides de taille similaire et de faible intensité sont également observés après purification de la protéine GST-NtLIM1 et clivage de la GST par la thrombine (Kawaoka *et al.*, 2000). Ces polypeptides pourraient aussi correspondre à des formes dégradées de la protéine. Par ailleurs, la migration observée sur gel SDS-PAGE de la protéine 6His-PtaGLIM1a à une taille supérieure (29 kDa) de celle prédite (21,9 kDa) n'est pas étonnante d'après les résultats obtenus par d'autres équipes. En effet, l'utilisation des anticorps A14 et A16 dirigés contre des peptides P14 et P16 respectivement situés dans les domaines C-terminal et inter-LIM de la protéine HaWLIM1, montre que cette protéine produite par *E. coli* migre vers 27 kDa (Brière *et al.*, 2003). De même, la protéine NtLIM1 de tabac purifiée par affinité est détectée vers 24 kDa après migration SDS-PAGE (Kawaoka *et al.*, 2000). Ainsi, ce polypeptide PtaGLIM1a détecté vers 29 kDa correspondrait donc à la taille attendue de la protéine produite par la bactérie. D'autre part, dans le pollen de tournesol, la protéine HaPLIM1 est aussi détectée vers 30 kDa par des anticorps monoclonaux

anti-PLIM1 (Baltz *et al.*, 1999). De plus, sur des extraits de protoplastes de tournesol, la protéine HaWLIM1 est observée vers 27 kDa avec les anticorps A14 ou A16. Ainsi, dans la plante, les protéines à domaine LIM sont probablement modifiées après leur traduction par des mécanismes encore inconnus. Il a en effet été suggéré que des modifications post-traductionnelles de type phosphorylation et glycosylation influent sur la migration de la protéine HaWLIM1 (Brière *et al.*, 2003). Par ailleurs, un polypeptide de taille plus élevée vers 50 kDa est aussi observé après purification de la protéine 6His-PtaGLIM1a. N'étant pas produit par la bactérie (Fig. 60), l'hypothèse la plus vraisemblable expliquant la présence de ce polypeptide de 50 kDa est une dimérisation de la protéine recombinante lors des étapes de purification. De plus, comme les protéines sont solubilisées dans un tampon contenant de l'urée et/ou du SDS et migrent en condition dénaturante, il est fort probable que la formation de dimère se fasse par l'intermédiaire de ponts disulfures entre les cystéines des doigts de zinc. De même, des résultats similaires ont été obtenus avec la protéine HaWLIM1 de tournesol purifiée par affinité. En effet, les anticorps dirigés contre la protéine GST-HaWLIM1 détectent principalement deux polypeptides : l'un vers 27 kDa et l'autre d'un poids moléculaire plus élevé vers 52 kDa (Brière *et al.*, 2003). Toutefois, la protéine NtWLIM1 de tabac purifiée par affinité ne migre pas à des tailles supérieures à 24 kDa (Kawaoka *et al.*, 2000). Par ailleurs, des polypeptides de haut poids moléculaire sont également révélés sur des extraits de protéines totales de peuplier, vers 36 et 45 kDa avec les anticorps A16, et de 70 à 100 kDa avec les anticorps anti-P-GLIM1a (Fig. 62 et 64). De même, concernant les protéines HaPLIM1 et HaWLIM1, des modifications post-traductionnelles ressemblant à une dimérisation de protéines LIM ont déjà été observées sur des extraits protéiques de plante (Baltz *et al.*, 1999 ; Mundel *et al.*, 2000). En effet, il a été montré que dans la majorité des tissus de tournesol, la protéine HaWLIM1 est détectée par des anticorps polyclonaux anti-GST-HaWLIM1 vers 50 kDa (Mundel *et al.*, 2000). Le polypeptide HaWLIM1 de haut poids moléculaire pourrait correspondre à un dimère entre protéines à domaines LIM ou entre une protéine LIM et un autre partenaire. En Western blot, un polypeptide HaPLIM1 de 50 kDa est aussi observé (en plus de celui migrant vers 30 kDa) (Baltz *et al.*, 1999). De plus, si de la iodoacétamide, un agent fixant irréversiblement les groupements thiols, est ajoutée à l'extrait protéique avant la migration en SDS-PAGE, seul le polypeptide de 30 kDa est détecté, indiquant probablement que le polypeptide de 50 kDa est un dimère de protéines. D'autre part, une homodimérisation de la protéine NtWLIM1 pourrait expliquer le fait qu'à saturation, deux moles de protéine NtWLIM1 s'associent avec une mole d'actine (Thomas *et al.*, 2006). Enfin, chez les animaux, il a été montré que les protéines

CRP1 et CRP3 sont capables de former des homodimères par l'interaction de leurs domaines LIM (Feuerstein *et al.*, 1994 ; Arber & Caroni, 1996). Cette formation d'homo ou hétérodimère supporte l'idée que les protéines à domaine LIM de plantes participeraient à des complexes moléculaires, de la même manière que les protéines CRP2 et CRP3 animales (Kong *et al.*, 1997 ; Louis *et al.*, 1997 ; Chang *et al.*, 2003). De plus, l'homodimérisation des protéines LIM pourrait aussi être un système de régulation modifiant leur activité biochimique. Il est donc fort probable que la plupart des protéines à domaine LIM de plantes puissent former des dimères *in vitro* et *in vivo*. Ce peut-être soit des homodimères entre même isoforme soit des hétérodimères entre différentes isoformes de protéines LIM ou entre des protéines LIM et d'autres protéines non apparentées.

#### 3.2. Les anticorps A16 détectent plusieurs protéines du groupe LIM1 chez le peuplier

L'utilisation des anticorps A16 nous donne des résultats intéressants mais difficilement interprétables car les six protéines LIM1 de peuplier peuvent être reconnues. En effet, un grand nombre de polypeptides de différentes tailles (20, 29, 36 et 45 kDa) sont détectés avec ces anticorps suggérant une reconnaissance non spécifique de plusieurs protéines à domaine LIM (Fig. 62). De plus, vers 29 kDa, ce sont deux bandes apparemment très proches entre elles qui sont décelées par les anticorps A16. Toutefois, de par les spécificités d'expression tissulaire des gènes *LIM* de peuplier, nous pouvons soulever des hypothèses sur la nature des polypeptides reconnus par ces anticorps. La détection du polypeptide de 20 kDa est inattendue car aucune analyse Western-blot précédente n'a détecté un polypeptide LIM de si petite taille (Baltz *et al.*, 1999 ; Kawaoka *et al.*, 2000 ; Mundel *et al.*, 2000 ; Brière *et al.*, 2003). De plus, la protéine HaWLIM1 est détectée à 27 kDa avec ces anticorps A16. Or, malgré sa petite taille, ce polypeptide de 20 kDa pourrait peut-être correspondre aux protéines du groupe LIM1 dont les tailles prédites sont comprises entre 21,4 kDa pour Pta $\beta$ LIM1a et b et 22 kDa pour PtaGLIM1a et b et PtaWLIM1a et b (Tableau 9). En effet, à l'inverse des autres protéines LIM de plante, ce polypeptide ne serait peut-être pas modifié par des phosphorylations ou glycosylations et migrerait à une taille correspondant approximativement à celle prédite. Ou bien, ce polypeptide de 20 kDa pourrait résulter d'autres types de modifications post-traductionnelles tel que des formes de clivage touchant spécifiquement cette protéine. Par ailleurs, ce polypeptide pourrait aussi bien être dégradé lors de l'extraction de protéine. Enfin, il est possible que ces anticorps A16, reconnaissant le

peptide P16 situé dans la région inter-LIM, détectent d'autres protéines non homologues aux protéines à domaine LIM. D'autre part, il est tout de même intéressant de constater l'abondance de ce polypeptide de 20 kDa uniquement dans le phloème et l'épiderme de la racine et l'écorce du tronc (Fig. 62). Lorsque nous comparons ce profil d'expression avec les résultats de RT-PCR, il est important de se rappeler que les gènes *PtaWLIM1a* et *PtaβLIM1a* exprimés de manière ubiquitaire, sont les seuls gènes fortement exprimés dans ces deux tissus (Fig. 48). De plus, les protéines HaWLIM1 et NtWLIM1 migrent à des tailles supérieures ou égales à 24 kDa (Kawaoka *et al.*, 2000 ; Mundel *et al.*, 2000). Ainsi, ce polypeptide de 20 kDa pourrait par déduction correspondre à la protéine PtaβLIM1a. Toutefois, nous ne pouvons exclure des artefacts de migration SDS-PAGE, et les protéines WLIM1 peuvent migrer à des tailles différentes en fonction de l'espèce étudiée. Enfin, il a été montré que la protéine HaWLIM1 est exprimée dans les cellules du phloème de tournesol (Mundel *et al.*, 2000).

Les deux polypeptides très proches vers 29 kDa pourraient correspondre à la taille attendue de la protéine PtaGLIM1a d'après les expériences de production de protéines recombinantes. Mais ces deux bandes pourraient aussi correspondre à des polypeptides d'autres protéines du groupe LIM1 ayant des tailles prédites voisines de celle de PtaGLIM1a (de 21,4 à 22 kDa). De plus, les polypeptides HaWLIM1 de tournesol sont également détectés à des tailles de 27 ou 34 kDa (Mundel *et al.*, 2000 ; Brière *et al.*, 2003). Etant donné que les gènes *PtaGLIM1a* et *b*, *PtaWLIM1b* et *PtaβLIM1b* ne sont pas exprimés dans le limbe et très faiblement dans les vaisseaux des feuilles (Fig. 48), le fort signal observé à 29 kDa dans ces tissus ne peut vraisemblablement correspondre qu'aux protéines PtaWLIM1a ou PtaβLIM1a. En effet, nous avons vu que ces deux gènes sont exprimés de manière ubiquitaire chez le peuplier. Par ailleurs, les polypeptides de 29 kDa ne sont pas détectés dans l'écorce du tronc et le phloème et l'épiderme de la racine, ces deux tissus ne contenant que le polypeptide LIM1 de 20 kDa (Fig. 62). Ainsi, nous pouvons imaginer que les protéines PtaWLIM1a ou PtaβLIM1a subissent des modifications post-traductionnelles particulières conduisant à un clivage de la protéine spécifiquement dans l'écorce/épiderme et le phloème. Par contre, la quantité élevée de ce polypeptide de 29 kDa dans les xylèmes matures de bois normal, opposé ou tendu suggère également une reconnaissance de la protéine PtaGLIM1a. En effet, les transcrits *PtaGLIM1a* s'accumulent fortement dans le xylème mature. De plus, le fort signal observé dans le xylème mature de bois tendu pourrait correspondre à la présence cumulée des protéines PtaWLIM1b et PtaGLIM1a et b. Enfin, le polypeptide de 36 kDa est détecté dans le xylème mature, spécialement dans celui de bois tendu, alors que le signal correspondant au



polypeptide de 45 kDa est plus fort dans les feuilles. Dans l'hypothèse où ces protéines formeraient des hétéro ou homodimères, ces deux types de polypeptides de poids moléculaire élevé indiqueraient une association avec différentes protéines partenaires au niveau des feuilles et du xylème mature.

#### 3.3. *Les anticorps anti-P-GLIM1a confirment l'induction de l'expression du gène PtaGLIM1a dans le bois de tension*

Les résultats d'immunodétection avec les anticorps anti-P-GLIM1a dans les différents tissus de peuplier, nous indiquent que la protéine PtaGLIM1a est bien exprimée dans les tissus vasculaires du tronc et de la racine, du cambium jusqu'au xylème mature (Fig. 64). La protéine est aussi faiblement détectée dans les pétioles, mais elle n'est pas décelée dans les vaisseaux, le limbe, les organes reproducteurs et les capsules contenant les graines et les fibres de coton. Ces résultats, ainsi que la forte accumulation de la protéine dans le xylème mature de bois tendu, sont en accord avec les résultats d'expression par RT-PCR semi quantitative du gène *PtaGLIM1a*. D'autre part, nous observons des polypeptides de masse très élevée entre 70 et 100 kDa et ayant des migrations diffuses. De plus, des expériences préliminaires de Western blot utilisant les anticorps dirigés contre la protéine recombinante 6His-PtaGLIM1a révèlent aussi des polypeptides de haut poids moléculaire à migration diffuse et étalée (Boizot N., communication personnelle). Ces résultats suggèrent fortement l'existence de modifications post-traductionnelles affectant la structure de la protéine PtaGLIM1a. Plus particulièrement, le caractère diffus de cette migration permet de penser qu'elle pourrait résulter de la glycosylation de la protéine. En effet, ce type de migration est observé pour les protéines AGP6 de pin et PopFLA1 et 8 de peuplier (Zhang *et al.*, 2003 ; Takeuchi *et al.*, 2007), des protéines à arabinogalactane connues pour être fortement glycosylées (Showalter, 2001). De plus, il existe deux sites putatifs de N-glycosylation à forte probabilité d'occurrence dans les régions inter-LIM et C-terminal. De même, 25 sites putatifs de O-glycosylation sont également présents mais leurs prédictions ont une probabilité plus faible, excepté un site positionné dans la région en C-terminal. Alternativement, cette protéine PtaGLIM1a pourrait interagir et former un dimère avec une protéine fortement glycosylée. Toutefois, il est étonnant de ne pas observer de migration diffuse des protéines LIM1 après révélation avec les anticorps A16. Ainsi, ces glycosylations seraient peut-être spécifiques à la protéine PtaGLIM1a, et les anticorps A16 pourraient reconnaître plus particulièrement les autres protéines du groupe LIM1. Ou alors, il est possible que ces glycosylations (ou

dimérisations) masquent la région antigénique reconnue par les anticorps A16. De plus, ce type de migration très diffuse n'a jamais été constaté pour les protéines WLIM1 et PLIM1 de plante (Baltz *et al.*, 1999 ; Mundel *et al.*, 2000 ; Kawaoka *et al.*, 2006). Par ailleurs, il est intéressant d'observer une migration différentielle des polypeptides de haut poids moléculaire et à migration diffuse en fonction de la différenciation du xylème. En effet, dans la zone cambiale, le polypeptide de 70 kDa prédomine, alors que dans le xylème mature celui-ci semble être remplacé par un polypeptide de taille supérieure vers 100 kDa (Fig. 64). De même, chez le tournesol, la taille de la protéine HaWLIM1 révélée en Western blot varie en fonction de l'âge des ovaires, avec un polypeptide de 34 kDa détecté dans les jeunes ovaires jusqu'à l'anthèse, et un polypeptide de 50 kDa principalement exprimé dans les ovaires au moment de l'anthèse et après pollinisation (Mundel *et al.*, 2000). Le résultat obtenu chez le peuplier peut être expliqué par des variations dans les modifications post-traductionnelles de la protéine PtaGLIM1a entre les cellules indifférenciées du cambium, les cellules de xylème en croissance et formant de la paroi primaire, et celles plus matures formant de la paroi secondaire. Ces modifications post-traductionnelles pourraient par exemple influencer la localisation subcellulaire de la protéine. En effet, il a été montré que les protéines CRP2 et CRP3 sont localisées dans le noyau des cellules indifférenciées, et à la fois dans le noyau et le cytoplasme des cellules musculaires (Arber *et al.*, 1994 ; Arber & Caroni, 1996 ; Chang *et al.*, 2003). Enfin, dans le xylème mature de bois tendu, deux polypeptides additionnels sont détectés vers 27 et 70 kDa. Les protéines PtaGLIM1a subiraient alors des modifications post-traductionnelles particulières dans le bois de tension. De même, nous pouvons aussi concevoir des interactions avec d'autres protéines partenaires spécifiquement exprimées dans le xylème mature de bois tendu. Ainsi, la fonction de cette protéine serait probablement différente en fonction de l'état de différenciation du xylème secondaire et du type de contraintes mécaniques perçues par la cellule.

## 4 Conclusions

En conclusion, malgré les difficultés rencontrées lors de la production et la purification de la protéine recombinante GST-PtaGLIM1a de par son association avec la protéine chaperonne GroEL et son insolubilité, l'utilisation de l'étiquette 6His nous a permis de contourner ce problème. En effet, nous avons pu produire en très grande quantité la protéine recombinante 6His-PtaGLIM1a sous forme de corps d'inclusion, puis celle-ci a été purifiée par affinité sur une résine de Cobalt et ensuite resolubilisée dans une solution ne contenant pas de détergent ni d'agent dénaturant. Deux lapins ont été immunisés avec cette protéine purifiée pour la production d'anticorps spécifiquement dirigés contre la protéine 6His-PtaGLIM1a. En parallèle, deux autres lapins ont été inoculés avec le peptide de synthèse P-GLIM1a situé dans la région C-terminale de la protéine, et choisi de manière à ce que les anticorps produits puissent différencier la protéine PtaGLIM1a de son duplicat la protéine PtaGLIM1b. Ces anticorps reconnaissent bien la protéine recombinante 6His-PtaGLIM1a produite par la bactérie et purifiée. Mais surtout, les anticorps anti-P-GLIM1a détectent la protéine PtaGLIM1a exprimée chez le peuplier. Les résultats de Western blot confirment au niveau protéique, l'expression spécifique du gène *PtaGLIM1a* dans les tissus vasculaires de l'arbre, et montrent une forte accumulation de la protéine dans le bois. Enfin, au niveau du xylème mature, la quantité de protéine est plus forte dans le bois de tension que dans le bois opposé, montrant une probable implication lors de la formation des fibres gélatineuses de bois tendu. Ces anticorps pourront ensuite servir pour d'autres applications, notamment l'immunodétection de la protéine sur des coupes histologiques de bois pour déterminer dans quel type cellulaire (fibres, rayons et vaisseaux) du xylème mature est exprimée la protéine PtaGLIM1a. Par ailleurs, ces anticorps pourront être utilisés pour des expériences d'immunoprécipitation de chromatine (ChIP) développées au laboratoire et pour lesquelles j'ai mis au point la première étape de purification de noyaux.

### Partie 3. La protéine PtaGLIM1a

## Discussion générale

### 1. Les protéines LIM de plantes ont une diversité structurale différente de celles des autres eucaryotes

Alors que chez les animaux les domaines LIM peuvent être en multiples copies au sein de la protéine et/ou être associés à d'autres domaines tels que des homéodomaines ou des domaines kinase, les protéines à domaine LIM de plantes sont en majorité apparentées aux protéines CRP animales et contiennent uniquement deux domaines LIM. De plus, au sein des protéines LIM de plantes, ces domaines LIM ne sont pas associés à d'autres motifs ayant une homologie avec des domaines connus. Néanmoins, il existe chez les plantes un faible nombre de protéines ne contenant qu'un seul domaine LIM associé à d'autres domaines tels que des motifs d'interaction à l'ubiquitine ou des métallopeptidase à zinc. Ainsi, comparées aux animaux, les plantes contiennent moins de protéines à domaine LIM, et celles-ci se sont peu diversifiées. Toutefois, si l'on compare la famille de protéines CRP animale à celle de protéines à domaine LIM de plante, cette dernière comprend plus de protéines. En effet, la famille CRP chez l'Homme ne contient que quatre protéines (CRP1, CRP2, CRP3 et TLP) (Weiskirchen & Günther, 2003), alors que les plantes peuvent contenir jusqu'à douze protéines à domaine LIM, comme c'est le cas chez le peuplier. Par ailleurs, ce domaine n'existe pas dans les protéines procaryotes, la levure ne possède que trois protéines à domaine LIM et les champignons ne contiennent qu'un nombre limité de protéines à domaine LIM (Kadmas & Beckerle, 2004). Ce domaine LIM semble donc important pour le fonctionnement des organismes pluricellulaires, et la diversification des protéines à domaine LIM chez les animaux semble être corrélée avec l'apparition d'organes ayant des rôles distincts dans l'organisme. Ainsi, les protéines CRP et LIM-HD se sont respectivement spécialisées dans des fonctions musculaires et neuronales chez les mammifères, et chez les plantes certaines protéines à domaines LIM ont peut-être acquis des fonctions impliquées dans le développement vasculaire ou l'élongation du tube pollinique. Comme pour les gènes LIM de plantes, les gènes CRP ont également acquis des spécificités d'expression tissulaire comme par exemple les muscles lisses vasculaires (CRP2) et les muscles cardiaques (CRP3) (Louis *et al.*, 1997). Dans, le génome de la mousse *Physcomitrella patens* récemment séquencé il n'y aurait que deux modèles de gène codant pour des protéines à domaine LIM très similaires entre elles. Ainsi, chez les plantes inférieures, le nombre de gènes LIM serait

probablement très faible comparé aux plantes supérieures. Chez la mousse, ces deux gènes pourraient assurer des fonctions cellulaires essentielles pour l'organisme. La famille de protéines à domaine LIM de plantes se serait ensuite considérablement agrandie et diversifiée avec le développement des plantes vasculaires et l'apparition de tissus spécialisés assurant de nouvelles fonctions liées, soit à la reproduction sexuée comme le pollen et les fibres de coton, soit au port de la plante avec les fibres de bois de tension.

### 2. *Les spécificités d'expression tissulaire des gènes LIM donnent des indications sur leurs fonctions potentielles lors du développement de la plante*

Les gènes codant pour des protéines LIM de plantes sont soit exprimés dans toute la plante et sont alors probablement essentiels dans des fonctions cellulaires de base (*WLIM1* et *WLIM2*), ou bien ont acquis une expression préférentielle dans certains tissus comme le pollen (*PLIM1*, *PLIM2* et  $\delta$ *LIM2*), les fibres vasculaires (*PtGLIM1a*, *PtGLIM1b* et *PtWLIM1b*), ou les fibres de coton (*GhaLIM1*). Les gènes *PLIM1* et *PLIM2* ne semblent pas être issus d'un ancêtre commun mais de deux processus évolutifs indépendants au sein des grands groupes LIM1 et LIM2. Ainsi, le caractère dérivé « spécificité pollinique » regroupant ces gènes *PLIM1* et *PLIM2* serait plutôt une analogie (homoplasie) d'expression résultant d'une convergence évolutive. Au contraire, le caractère dérivé « expression constitutive » commun aux gènes *WLIM1* et *WLIM2* pourrait être hérité d'un ancêtre commun présent chez la mousse *Physcomitrella patens*, et nous aurions alors peut-être une homologie d'expression.

Pour la première fois, l'analyse de l'expression des gènes *LIM* chez le peuplier montre une possible implication de certaines protéines à domaine LIM (*PtGLIM1a* et *b*, *PtWLIM1a* et *b* et *PtPLIM2a*) dans le développement vasculaire de la plante. De même, les gènes *AtWLIM1*, *NtWLIM1* et *EcWLIM1* sont plus fortement exprimés dans les tiges que dans d'autres parties de la plante (Eliasson *et al.*, 2000 ; Kawaoka *et al.*, 2000 ; Kawaoka *et al.*, 2006), suggérant également une fonction dans le développement des tissus vasculaires. Toutefois, la localisation cellulaire précise de ces protéines n'a pas encore été testée. Chez le pin, le gène *PitLIM1b* est aussi préférentiellement exprimé dans le xylème (Allona *et al.*, 1998). Ainsi, nous pouvons supposer que le gène *LIM1* ancestral avait également une expression préférentielle dans les tissus vasculaires. Les tissus conducteurs de la plante sont présents dans une grande variété d'organes, et ceci expliquerait peut-être le caractère constitutif de l'expression des gènes *WLIM1* des angiospermes et *LIM1* de conifères. De même, nous pouvons envisager que le gène  $\beta$ *LIM1b* soit spécifique des tissus vasculaires au

niveau des organes reproducteurs. Les tissus vasculaires sont importants pour la physiologie des arbres car ils permettent à la fois le transport d'eau sur de longues distances et le support du poids de la plante (Mellerowicz *et al.*, 2001 ; Ye, 2002). Ces propriétés pourraient expliquer l'expansion de cette famille LIM chez le peuplier avec l'apparition de gènes supplémentaires (*PtGLIM1a* et *b*, *PtWLIM1b*) ayant une fonction putative dans la différenciation des vaisseaux, des rayons ou des fibres de bois. Bien que les profils d'expression suggèrent plutôt un rôle lors de l'expansion cellulaire ou de la synthèse de parois primaire et secondaire, une analyse d'expression plus précise au niveau cellulaire est nécessaire afin de comprendre l'implication de ces gènes lors des différentes étapes de la xylogénèse, de la division des initiales fusiformes jusqu'à la mort cellulaire programmée. Par ailleurs, l'expression du gène *SoWLIM1b* de canne à sucre est probablement très forte dans la zone de transition tige/racine où les tissus vasculaires changent d'organisation. Il a été montré chez *Arabidopsis* une forte accumulation de composés flavonoïdes dans cette zone (Murphy & Taiz, 1999 ; Peer *et al.*, 2001). De plus, il est important de noter que l'expression du gène *PGPS/D1* de pétunia dans le pollen est induite par le kœmpférol, un composé de la voie de biosynthèse des flavonoïdes (Guyon *et al.*, 2000). Comme pour le gène *PGPS/D1*, l'expression du gène *SoWLIM1b* pourrait être activée par les flavonoïdes. Enfin, ces données suggèrent fortement un rôle des flavonoïdes comme molécule signal régulant à la fois la différenciation des tissus vasculaires et la formation du tube pollinique.

Les gènes *PLIM2*, *δLIM2* et *PLIM1* sont fortement exprimés dans le pollen de façon similaire et sont probablement impliqués dans l'élongation du tube pollinique. Toutefois, les gènes *PLIM2* et *δLIM2* sont retrouvés dans toutes les plantes supérieures alors que les gènes *PLIM1* sont spécifiques des astéridées. Ces espèces ont sûrement acquis avec l'apparition des protéines PLIM1 une fonction particulière conférant un avantage sélectif aux grains de pollen. Par ailleurs, les gènes *PLIM2* sont en général en multiple copie chez les plantes supérieures avec par exemple trois gènes *AtPLIM2a*, *b* et *c* chez *Arabidopsis* et quatre gènes *PtPLIM2a* et *b* et *PtδLIM2a* et *b* chez le peuplier qui sont tous fortement exprimés dans le pollen. Ainsi, ces gènes présentent sûrement une redondance fonctionnelle lors de la croissance du tube pollinique, et cette particularité sera à prendre en compte lors de l'étude de leurs fonctions. Chez les espèces de la classe des astéridées, les protéines PLIM1 sont phylogénétiquement très différentes entre elles surtout au niveau de la région C-terminale. Il en est de même pour les protéines PLIM2 et *δLIM2* de monocotylédones, eudicotylédones et d'astéridées exhibant des régions C-terminales très diversifiées. Cette diversité est peut-être la conséquence d'une adaptation évolutive aux différences morphologiques ou biochimiques existant entre ces

espèces au niveau des organes récepteurs femelles tels que le stigmate, le style, le micropyle et la cavité ovarienne. De même, ces différences pourraient éventuellement empêcher les croisements interspécifiques. Ainsi, les gènes *PLIM1* des familles d'*Asteraceae* et de *Solanaceae* répondent sûrement à des spécificités reproductives chez ces espèces ayant, faut il le remarquer, une morphologie des organes récepteurs femelles globalement différente de celle des rosidées. Enfin, il est important de noter qu'aucun gène *PLIM1* et *PLIM2* n'a été identifié chez les conifères. La pollinisation chez les conifères diffère nettement de celle des angiospermes. Tout d'abord, les ovules sont nus sans ovaire, et la fleur femelle ne possède pas de stigmate ou de pistil afin de sélectionner les grains de pollen pour leurs capacités à produire des tubes polliniques performants (Philippe *et al.*, 2006). A la place, la fleur contient un nucelle et un micropyle sur lesquels très peu de grains de pollen sont naturellement déposés. En conséquence la compétition pollinique est très faible chez les conifères. De plus, la vitesse de croissance du tube pollinique est beaucoup plus lente chez les gymnospermes que chez les angiospermes, et le temps entre la pollinisation et la fertilisation est plus important. Par exemple, la croissance du tube pollinique dure deux semaines chez le mélèze et le Douglas, de deux à quatre semaines chez les épicéas et une année chez les pins où la progression du tube pollinique est interrompue pendant l'hiver (Philippe *et al.*, 2006). Dans l'hypothèse où les protéines *PLIM1* et *PLIM2* aident à l'élongation du tube pollinique, il n'est alors pas étonnant que les conifères ne possèdent apparemment pas ces protéines qui ne leurs donneraient aucun avantage sélectif.

Chez le cotonnier, à l'exception des gènes *GrPLIM2b* et *GrPLIM2c*, tous les autres gènes comportent des ESTs issues de banques d'ADNc de fibres de coton, surtout le gène *GhaLIMI* du sous-groupe *FLIM1* apparemment préférentiellement exprimé dans les fibres de coton. Les fibres de coton dérivent de cellules épidermiques de l'ovaire et entourent la graine au cours de son développement. Ce sont des trichomes de graines unicellulaires et contrairement aux trichomes de feuilles, ils ne se ramifient jamais (Ruan, 2007). Comme les cellules du xylème, les fibres de coton subissent quatre stades de développement qui se chevauchent : initiation, élongation, synthèse de la paroi secondaire et maturation. La croissance rapide des fibres en élongation est couplée avec une synthèse constante et abondante de composants de la paroi cellulaire comme la cellulose qui représente environ 95 % du poids sec d'une fibre de coton. Contrairement aux cellules de bois, la paroi secondaire des fibres de coton ne contient pas de lignine. Ainsi, la paroi des fibres de coton peut être comparée à la couche G des fibres gélatineuses du xylème de par leurs forts contenus en cellulose. Toutefois, par rapport aux fibres G, l'élongation cellulaire est beaucoup plus



importante chez les fibres de coton, puisqu'elles peuvent atteindre jusqu'à 2 mm. Nous avons vu que les gènes *PtGLIM1a* et *b* spécifiques des tissus vasculaires et dont l'expression est induite dans le bois de tension, ne sont pas exprimés dans les fibres de coton de peuplier, alors qu'ils sont phylogénétiquement proches du gène *GhaLIM1* de cotonnier. Chez le peuplier, d'autres gènes (*PtWLIM1a*, *PtβLIM1a* et *b*, *PtPLIM2a* et *b*, et *PtδLIM2a* et *b*) semblent assurer une fonction dans le développement des fibres de coton. Au sein du sous-groupe FLIM1, seul le cotonnier et le peuplier sont des espèces formant des graines avec des poils cotonneux. En effet, le ricin et l'euphorbe de la famille des *Euphobiaceae* produisent des graines dépourvues de tels poils et qui ne semblent pas être disséminées par le vent. Ainsi, parmi les espèces identifiées dans le sous-groupe FLIM1, seul le cotonnier possède un gène ayant une fonction putative dans le développement des fibres de coton, et ce caractère semble donc apparu de façon indépendante au cours de l'évolution. Par contre, quelle que soit l'espèce, les gènes groupés dans le sous-groupe FLIM1 pourraient être tous impliqués dans la formation de fibres gélatineuses. En effet, le ricin et certaines espèces d'euphorbe et de cotonnier comme *Euphorbia tirucalli* et *Gossypium arboreum* sont des plantes arborescentes et ligneuses. Ces plantes pourraient donc sous l'effet de contraintes mécaniques développer des fibres G de xylème comme le peuplier. De plus, il a été montré que des plantules de ricin placées horizontalement forment des fibres gélatineuses au niveau des tissus primaires de l'hypocotyle (Kang & Soh, 1994). Il est toutefois surprenant de ne pas avoir identifié de gènes *FLIM1* chez l'eucalyptus alors que certaines espèces du genre développent des fibres G. De plus, les analyses transcriptomiques n'ont pas détecté de transcrits codant pour des protéines à domaine LIM et ayant une expression dérégulée dans le bois de tension d'eucalyptus (Paux *et al.*, 2005). Néanmoins, les gènes du sous-groupe FLIM1 pourraient par exemple être impliqués dans l'organisation du cytosquelette d'actine, l'inhibition de la synthèse de lignine ou l'activation de la synthèse de cellulose, des fonctions cellulaires à la fois importantes pour le développement des fibres de coton et celui des fibres gélatineuses du xylème.

Enfin, de manière intéressante, l'analyse des données d'expression microarray fournies par la base de données NASCArray indique que les gènes *AtWLIM1*, *AtWLIM2a*, et *AtPLIM2b* sont fortement exprimés dans la zone de différenciation des poils racinaires (Craigon *et al.*, 2004 ; Jones *et al.*, 2006). Par rapport aux plantes sauvages, l'expression de ces gènes est fortement diminuée d'un facteur deux chez le mutant « *root hair defective 2* » (*rhd2*) codant pour un homologue d'une NADPH oxydase produisant des peroxydes d'hydrogène (H<sub>2</sub>O<sub>2</sub>). La croissance du pollen et des poils racinaire nécessite la présence d'H<sub>2</sub>O<sub>2</sub> au niveau de l'extrémité de la cellule. Chez ce mutant, l'initiation de la formation des

poils racinaire est normale, mais ceux-ci sont courts car ils sont incapables d'avoir une croissance à leurs extrémités (Jones *et al.*, 2006). *AtWLIM1*, *AtWLIM2a*, et *AtPLIM2b* sont donc probablement des gènes également impliqués dans la morphogenèse et l'élongation des poils racinaires.

L'ensemble des résultats d'expression suggère un rôle plus général des protéines à domaine LIM de plante dans la croissance cellulaire anisotrope et peut-être isotrope. En effet, certains gènes sont exprimés dans le pollen en germination ou les poils racinaires ayant une croissance intrusive à l'extrémité de la cellule (Hepler *et al.*, 2001 ; Chen *et al.*, 2006), dans les fibres et/ou vaisseaux du xylème à croissance diffuse et/ou intrusive (Mellerowicz *et al.*, 2001 ; Chaffey *et al.*, 2002 ; Ye, 2002), et dans les fibres de coton à croissance plutôt diffuse (Ruan, 2007). De plus, certains gènes *LIM* pourraient intervenir dans l'élongation cellulaire au niveau de différents types de tissus et avoir ainsi des fonctions cellulaires similaires dans des processus physiologiques différents. En effet, les résultats d'analyses d'expression RT-PCR montrent que le gène *PtPLIM2a* est à la fois fortement exprimé dans le pollen, les fibres cotonneuses et le xylème en développement. Chez le peuplier, ce gène pourrait donc avoir un rôle important dans la croissance anisotrope des cellules, à la fois au niveau des tubes polliniques, des fibres de cotons et dans les fibres ou vaisseaux du xylème. Par ailleurs, les gènes *WLIM1* et *WLIM2* pourraient avoir une fonction plus globale dans l'expansion cellulaire isotrope et anisotrope. Lors de ces processus d'élongation cellulaire, il est fort probable que les protéines LIM aient au moins deux rôles dans l'organisation du cytosquelette d'actine et dans l'élaboration de la paroi cellulaire primaire et/ou secondaire.

### 3. Les protéines LIM : une double fonction cellulaire ?

#### 3.1. Des protéines se liant à l'actine

Les cellules subissant une forte élongation anisotrope nécessitent en général de longs câbles d'actine pour apporter les précurseurs pariétaux vers les sites de synthèse de la paroi cellulaire. Chez le tabac, la protéine NtWLIM1 stabilise et empaquette les câbles de filaments d'actine (Thomas *et al.*, 2006). Ainsi, les spécificités d'expression des gènes *LIM* dans différentes cellules ayant en commun une forte expansion anisotrope, confortent un rôle des protéines à domaine LIM de plante dans la formation et le maintien de câbles d'actine

facilitant le flux cytoplasmique. En effet, dans le bois, les initiales fusiformes possèdent d'épais filaments d'actine parallèles à l'axe longitudinal de la cellule probablement nécessaires à la synthèse de la paroi cellulaire aux extrémités des cellules (Chaffey *et al.*, 2002). De plus, lors de la croissance en diamètre des initiales fusiformes, certaines protéines LIM comme PtGLIM1b et PtWLIM1b, pourraient participer à la formation des câbles d'actine attachés aux ponctuations simples, participant ainsi à l'augmentation du branchement latéral du réseau de microfilaments d'actine dans les fibres et les vaisseaux (Chaffey *et al.*, 2002). Les protéines LIM de peuplier pourraient aussi en stabilisant les câbles d'actine intervenir dans la perception et le contrôle des contraintes mécaniques générées dans les cellules de xylème ou dans les tubes criblés du phloème. En effet, chez les animaux, les protéines CRP1 et CRP2 se fixent aux fibres de stress d'actine (Grubinger & Gimona, 2004 ; Tran *et al.*, 2005) dont la formation est induite par la tension des cellules, et la protéine CRP2 pourrait réguler la force contractile des cellules (Kim-Kaneyama *et al.*, 2005). D'autre part, les gènes *LIM* exprimés dans la zone cambiale de peuplier pourraient être impliqués dans la formation du phragmoplaste et de la plaque cellulaire. En effet, chez le tournesol, la protéine HaWLIM1 est localisée au niveau du phragmoplaste des cellules en division du cambium interfasciculaire (Mundel *et al.*, 2000). De plus, les initiales cambiales de peuplier se divisant par leur côté le plus long, le phragmoplaste se trouve étiré d'une extrémité à l'autre de la cellule (Larson, 1994 ; Mellerowicz *et al.*, 2001). Le réseau d'actine et peut-être les protéines LIM pourraient alors jouer un rôle particulier dans le transport vésiculaire des molécules nécessaires à la synthèse de la paroi cellulaire du phragmoplaste et de la plaque cellulaire.

Dans le xylème, les gènes *LIM* de peuplier sont peut-être exprimés dans d'autres cellules que les fibres, comme les vaisseaux et les cellules de rayon. Il a été montré que l'actine est fortement localisée au niveau des ponctuations simples, de contact et aréolées (Chaffey *et al.*, 2002). Les protéines LIM pourraient ainsi intervenir dans le développement de ces ponctuations, ou bien lors de la formation de bois de tension, jouer un rôle dans la transduction du signal mécanique entre les cellules en contrôlant l'ouverture de ces plasmodesmes. Bien que cette dernière hypothèse soit attrayante, il faut la prendre avec précaution car nos connaissances sont très limitées sur la nature du signal transmettant le stimulus mécanique (Baluska *et al.*, 2003 ; Telewski, 2006) et sur le système de contraction « actine-myosine » des ponctuations aréolées proposé par (Chaffey & Barlow, 2002). Comme je l'ai suggéré auparavant, les protéines LIM pourraient également participer dans les différentes cellules du xylème à la formation des câbles d'actine attachés à ces ponctuations afin de guider le transport de composés pariétaux vers ces sites de synthèse de paroi cellulaire.

Dans les cellules de rayon en fin de différenciation et matures, les protéines LIM peuvent participer à l'épaississement des câbles d'actine pour le transport de matériel à travers le phloème et le xylème par la voie symplasmique.

Chez les mammifères, les protéines CRP1 et CRP2 interagissent avec les fibres de stress d'actine (Grubinger & Gimona, 2004 ; Tran *et al.*, 2005), intervenant dans la contraction des cellules, et attachées aux adhésions focales par l'intermédiaire des intégrines et d'autres protéines se liant à l'actine. Toutefois, les adhésions focales au sens strict ne semblent pas exister chez les plantes, puisque les génomes de plantes ne possèdent pas de vrais homologues aux intégrines et  $\alpha$ -actinines animales (Baluska *et al.*, 2003), et les cellules végétales possèdent une paroi cellulaire. Les cellules végétales possèdent sûrement un autre système équivalent pour la perception du signal extracellulaire au niveau de la membrane plasmique et de la paroi cellulaire, et aussi pour le contrôle de la formation des câbles d'actines dans le cytoplasme et leurs attachements à la membrane plasmique. Ce système recruterait d'autres protéines spécifiques de plante telles que des protéines de type AGP, WAK, LIM et/ou CesA. De plus, dans les cellules musculaires, les protéines CRP2 et CR3 apparaissent essentielles dans la perception du stress mécanique (Knoll *et al.*, 2002 ; Kim-Kaneyama *et al.*, 2005). La découverte de l'induction d'un grand nombre de gène *LIM* de peuplier dans le bois de tension, un tissu subissant et générant de fortes contraintes mécaniques est intrigante si l'on compare l'évolution fonctionnelle des protéines CRP animale et LIM de plante. La fonction entre ces protéines LIM apparentées est probablement conservée entre les règnes animaux et végétaux (Thomas *et al.*, 2006 ; Thomas *et al.*, 2007). Il est ainsi probable que les protéines à domaine LIM de plante aient également un rôle dans la réponse des cellules au stress mécanique en régulant l'organisation du cytosquelette d'actine et la transcription des gènes. De plus, les fibres de xylème ne sont pas les seules cellules subissant des contraintes mécaniques. Les tubes polliniques et les poils racinaires pénétrant respectivement dans le stigate/pistil et la rhizosphère doivent également résister à des forces mécaniques importantes. Par ailleurs, les protéines LIM pourraient aussi intervenir, par l'intermédiaire du cytosquelette d'actine, dans la perception ou la réponse cellulaire à la pression de turgescence. En effet, dans ces cellules à croissance rapide la pression de turgescence peut augmenter fortement et la cellule doit réagir en augmentant l'élasticité de la paroi primaire. Enfin, un rôle direct des protéines LIM dans la perception du stimulus gravitationnel est envisageable. En effet, chez le peuplier, le stimulus gravitationnel serait probablement perçu au niveau des cellules endodermique de la tige comme pour les plantes herbacées (Blancaflor & Masson, 2003) ; et lors de l'inclinaison des arbres, l'expression des

gènes *PtWLIM1a*, *PtWLIM1b* et *Pt $\delta$ LIM2a* est induite dans l'écorce et le phloème des tiges. De plus, dans le cortex de la tige de tournesol, la protéine HaWLIM1 est associée aux amyloplastés (statolithes), des organites attachés aux filaments d'actine (Mundel *et al.*, 2000). Ainsi ces protéines LIM pourraient fonctionner dans la perception ou la transduction du signal gravitationnel en influençant la sédimentation des amyloplastés par le contrôle de l'empaquetage des filaments d'actine. En effet, il a été montré que l'application de latrunculine B réduit sévèrement la mobilité des amyloplastés et il a été suggéré que la sédimentation des amyloplastés est dépendante de l'actine filamenteuse (Palmieri & Kiss, 2005).

D'autre part, il a été montré que les protéines CRP ont de multiples partenaires dans la cellule et sont capables d'interagir avec l' $\alpha$ -actinine et la zyxine (Sadler *et al.*, 1992 ; Schmeichel & Beckerle, 1994 ; Louis *et al.*, 1997 ; Pomies *et al.*, 1997). Toutefois, il n'existe pas d'homologues à ces deux protéines chez les plantes. Il est fort probable que les protéines à domaine LIM de plante soient aussi capables de se lier à d'autres protéines régulant l'organisation du cytosquelette. Mais à l'heure actuelle, ces partenaires protéiques ne sont pas encore connus. Les protéines à domaine LIM chez les plantes sont probablement toutes capables de se lier à l'actine, comme cela a été montré pour la protéine *NtWLIM1* de tabac (Thomas *et al.*, 2006). De la même façon, la colocalisation de la protéine HaWLIM1 avec le réseau de microtubule suggère une association indirecte des protéines LIM au cytosquelette de microtubules (Brière *et al.*, 2003). Ces deux résultats mènent à l'hypothèse que ces protéines LIM pourraient intervenir dans l'interconnexion entre les filaments d'actine et les microtubules. Il a notamment été observé d'étroites relations entre ces deux éléments du cytosquelette dans les tubes polliniques chez le coquelicot (*Papaver rhoeas*) et le pin (*Picea abies*) (Justus *et al.*, 2004 ; Gossot & Geitmann, 2007). De même, il a été suggéré une coopération entre les microtubules et l'actine pendant la phase d'épaississement de la paroi secondaire des éléments de vaisseaux (Chaffey *et al.*, 2002). De plus, le degré de coopération entre ces deux éléments du cytosquelette est bien marqué dans les cellules du xylème notamment au niveau des ponctuations aréolées et de contact. Par ailleurs, l'hypothèse d'une interaction des protéines LIM avec le domaine à doigt de zinc de certaines cellulose synthases est attrayante, surtout pour les gènes *LIM* fortement exprimés dans les fibres de coton ou les fibres G de bois de tension, des cellules ayant une paroi à forte teneur en cellulose. Enfin, les analyses en Western blot nous indiquent que la protéine PtaGLIM1a forme probablement des homo ou hétérodimères, et subit différentes modifications post-traductionnelles en fonction de la différenciation du xylème et de la mise en tension des tiges. De façon similaire, en Western

blot, la quantité du peptide HaWLIM1 de 50 kDa augmente en fonction de la maturation des ovaires, alors que celle du peptide de 34 kDa diminue (Mundel *et al.*, 2000). Les différences de taille du peptide PtaGLIM1a de haut poids moléculaire pourraient être le résultat de son interaction avec différents partenaires protéiques au cours des étapes de la xylogénèse. De plus, le caractère diffus de cette migration suggère que ces partenaires protéiques sont peut-être glycosylés et donc seraient plutôt localisés au niveau des membranes plasmiques ou intracellulaires. Il a en effet été observé chez le tournesol une localisation de la protéine HaPLIM1 au niveau de la région corticale des cônes germinatifs des grains de pollen (Baltz *et al.*, 1999), et la protéine HaWLIM1 est souvent associée à des vésicules ou organelles intracellulaires (Mundel *et al.*, 2000).

Les protéines PLIM1 et PLIM2 ont sûrement acquis des fonctions spécifiques impliquées dans la formation du tube pollinique, alors que les protéines WLIM1 et WLIM2 auraient une fonction plus générale lors du développement de la plante. Il est important de remarquer que certains membres appartenant à d'autres familles de protéines se lient à l'actine tels que les familles profiline, formine, ADF/cofiline, villine, gelsoline/fragmine, et le complexe Arp2/3 ont aussi des fonctions importantes lors du développement du tube pollinique (Ren & Xiang, 2007). Prenons par exemple le cas de la famille profiline chez les plantes. Il est très intéressant de constater une évolution semblable dans l'organisation phylogénétique des familles LIM et profiline, ainsi que dans les spécificités d'expression de leurs gènes. En effet, parmi les cinq isoformes de profiline chez *Arabidopsis thaliana* (AtPRF1 à 5), les gènes *AtPRF1*, 2 et 3 sont exprimés dans les tissus végétatifs et les gènes *AtPRF4* et 5 sont spécifiquement exprimés dans le pollen au stade tardif de son développement et dans les grains de pollen en germination (Christensen *et al.*, 1996 ; Huang *et al.*, 1996 ; Kandasamy *et al.*, 2002). Cette spécificité d'expression pollinique est d'autant plus intéressante qu'elle est, comme pour les protéines à domaine LIM, liée à un groupe monophylétique distinct au sein de cette famille (Huang *et al.*, 1996). De plus, les gènes codant pour des profilines végétatives AtPRF1 et 2 sont préférentiellement exprimés dans les paquets vasculaires (Christensen *et al.*, 1996 ; Ramachandran *et al.*, 2000), de la même manière que les protéines WLIM1 ou FLIM1 de peuplier. Cette similarité de profils d'expression des gènes appartenant à ces deux familles de protéines capables de se lier à l'actine met en évidence de façon générale l'importance de ce type de protéines à la fois dans le développement des tissus vasculaires et dans la croissance du tube pollinique. Par ailleurs, la protéine AtPRF1 est capable de compléter les mutants *profiline* des levures *Saccharomyces cerevisiae* et *S. Pombe* montrant que les profilines de plantes sont

fonctionnellement similaires à celles de levure, et ce malgré la faible homologie de séquence entre elles (Christensen *et al.*, 1996). De la même manière, bien qu'ayant une homologie relativement faible avec les protéines CRP animale, les protéines à domaines LIM de plante semblent avoir une fonction similaire dans la fixation et l'empaquetage des filaments d'actine (Grubinger & Gimona, 2004 ; Tran *et al.*, 2005 ; Thomas *et al.*, 2006 ; Thomas *et al.*, 2007). Cette similarité de fonction entre ces protéines LIM de plante et CRP animale souligne une fois de plus le fort degré de conservation observé dans le système contrôlant la balance entre l'actine filamentueuse et globulaire chez les cellules eucaryotes (Cvrcková *et al.*, 2004).

### 3.2. Facteurs de transcription régulant des gènes de la voie de biosynthèse des lignines

Alors que chez les animaux les protéines CRP ne semblent pas capables d'interagir directement avec l'ADN et ont un rôle de cofacteur de transcription (Kong *et al.*, 1997 ; Chang *et al.*, 2003), la protéine NtLIM1 de tabac est capable de se fixer à l'ADN et d'activer la transcription des gènes (Kawaoka & Ebinuma, 2001). De plus, lorsque l'expression du gène *NtWLIM1* est réprimée, la teneur en lignine dans la plante diminue sensiblement, ainsi que l'expression des gènes *PAL*, *4CL* et *CAD*. Cependant, certains résultats obtenus par ce travail ne concordent pas avec une fonction de ces protéines LIM dans l'activation de la biosynthèse de la lignine, et plusieurs études ont également remis en cause cette propriété (Laitinen *et al.*, 2005 ; Rogers *et al.*, 2005 ; Andersson-Gunneras *et al.*, 2006). Par exemple, dans la tige d'*Arabidopsis* poussant à l'obscurité et en comparaison avec les plantes témoins (sans ajout de saccharose), l'expression du gène *AtWLIM1* est réprimée par l'ajout de saccharose, un traitement favorisant la biosynthèse des lignines (Rogers *et al.*, 2005). De plus, l'expression du gène *AtWLIM1* est aussi réduite chez les mutants *eli1*, *det3* et *pom1* présentant une lignification ectopique des tiges, et il a été suggéré que *AtWLIM1* soit un répresseur de la biosynthèse des lignines. Toutefois, si l'on considère que les gènes orthologues ont une fonction conservée, cette hypothèse n'est pas compatible avec les résultats obtenus sur la protéine NtLIM1 de tabac, un facteur de transcription activateur favorisant la lignification des tiges (Kawaoka *et al.*, 2000). Alternativement, la surproduction de lignine chez ces mutants *eli1*, *det3* et *pom1* pourrait induire une répression de l'expression du gène *AtWLIM1* par une boucle de rétrocontrôle négative, afin de limiter une suraccumulation excessive de lignine dans la plante.

Par ailleurs, beaucoup de gènes *LIM* sont préférentiellement exprimés dans les cellules en élongation formant de la paroi primaire. Cette étape d'expansion cellulaire est incompatible avec une production de lignine, un composé qui rigidifie et imperméabilise les parois secondaires, celles-ci n'étant ainsi plus extensibles. Il est alors étonnant de constater au sein de cette famille de protéines à domaine LIM, l'existence des gènes *PLIM1* et *PLIM2* préférentiellement exprimés dans le pollen et les tubes polliniques. Ces derniers possèdent une paroi cellulaire composée essentiellement d'une gaine intérieure de callose, et d'une couche extérieure de cellulose, d'hémicellulose et de pectine ; et cette paroi ne semble pas contenir de lignine (Southworth, 1973 ; Taylor & Hepler, 1997). En effet, bien que les activités de la PAL et de peroxydases augmentent durant la germination de grains de pollen de maïs suggérant la présence de polymères de lignine dans la paroi cellulaire de pollen (Liu & Ger, 1997), aucune étude n'a mis en évidence la présence de lignine dans le pollen. Seules les cellules du tapetum de l'anthère contiennent de la lignine dans leur paroi. Il a ainsi été suggéré que les gènes *PLIM1* et *PLIM2* de *Gerbera hybrida* spécifiquement exprimés dans les étamines aient une fonction non liée à la biosynthèse des lignines (Laitinen *et al.*, 2005). Par contre, il a été montré que certains flavonols s'accumulent tardivement au cours du développement du pollen et sont importants pour sa germination (Mo *et al.*, 1992 ; Taylor & Hepler, 1997). Les résultats montrant que l'expression du gène *PGPS/D1* de pétunia est induite dans le pollen par les flavonoïdes est intéressante en regard de la fonction présumée de facteur de transcription régulant l'expression de gènes appartenant à la voie des phénylpropanoïdes (Guyon *et al.*, 2000 ; Kawaoka *et al.*, 2000). Les deux voies métaboliques des monolignols et des flavonoïdes sont étroitement liées, car au sein des familles de gènes de la voie générale des phénylpropanoïdes (PAL, C4H, 4CL), les isoenzymes sont impliquées soit dans la voie de biosynthèse des flavonoïdes soit dans celle des lignines (Boerjan *et al.*, 2003 ; Tsai *et al.*, 2006). Par ailleurs, il a été montré que la suppression de l'expression du gène *HCT* chez *Arabidopsis* provoque une répression de la biosynthèse de lignine conduisant à un changement du flux métabolique vers les flavonoïdes via la chalcone synthase (Besseau *et al.*, 2007). De plus, chez ces plantes transgéniques accumulant des flavonoïdes, la morphologie des éléments de trachéides est altérée et le transport d'auxine est inhibé. Nous pouvons alors supposer que dans le pollen, ces protéines PLIM1 pourraient réprimer fortement l'expression des gènes spécifiques de la voie de biosynthèse de la lignine, ce qui amènerait à une accumulation de flavonoïdes nécessaires à la germination du pollen et la croissance du tube pollinique. Alternativement, nous pouvons suggérer un autre modèle où les protéines PLIM1 pourraient activer dans le pollen l'expression des gènes de la voie de biosynthèse des



flavonoïdes plutôt que celle des lignines. Et par une boucle de rétrocontrôle positive, les flavonoïdes activeraient à leurs tours l'expression des gènes *PLIM1* pour avoir une quantité constante de flavonoïdes dans le tube pollinique.

Un grand nombre de gène *LIM* de peuplier ont une expression induite dans le bois de tension ou sont fortement exprimés dans les fibres de coton. Comme pour les tubes polliniques, les fibres de coton semblent totalement dépourvues de lignine (Ruan, 2007), et une fonction des ces gènes *LIM* dans l'activation de la biosynthèse des lignine est alors exclue. Un rôle du gène *PttGLIM1b* dans l'activation de l'expression des gènes *PAL*, *CAD* et *4CL* favorisant la synthèse de lignine dans le bois de tension a également été mis en doute (Andersson-Gunneras *et al.*, 2006). En effet, les fibres G de bois de tension possèdent une couche gélatineuse essentiellement constituée de cellulose (Timell, 1969) et presque dépourvue de lignine (Joseleau *et al.*, 2004 ; Pilate *et al.*, 2004 ; Gierlinger & Schwanninger, 2006). Toutefois, les couches S1 et S2 des fibres G, bien que plus fines par rapport à celles de bois normal, apparaissent normalement lignifiées (Timell, 1969 ; Joseleau *et al.*, 2004). Ainsi, dans le bois de tension, les gènes *PtGLIM1a* et *b*, *PtWLIM1b* et *PtPLIM2a* pourraient éventuellement participer à la lignification de ces couches, au moins durant les étapes précoces de la formation de la paroi secondaire. De plus, dans les tiges de tabacs transgéniques antisens *Ntlim1*, la composition en unité guaiacil et syringil semble modifiée, et la protéine NtWLIM1 pourrait favoriser l'augmentation du rapport d'unité S/G dans la lignine (Kawaoka *et al.*, 2000). De même, dans le bois de tension, la composition en unité S et G varie et le rapport S/G semble augmenter (Aoyama *et al.*, 2001 ; Joseleau *et al.*, 2004). Les fibres et rayons contiennent aussi une plus forte proportion d'unités S que d'unités G, alors que les vaisseaux sont enrichis en lignine de type G (Boerjan *et al.*, 2003). Nous pouvons alors penser que les protéines LIM seraient plus impliquées dans la lignification des fibres et rayons du xylème que dans celle des éléments de vaisseaux. De plus, dans les tiges plus âgées la teneur en unité S augmente (Meyer *et al.*, 1998). Si les gènes *PtGLIM1a* et *b* sont effectivement impliqués dans la synthèse d'unité S de lignine, ceci concorderait avec leur forte expression dans le xylème secondaire des tiges de peuplier. Par contre, il a été suggéré que la présence de l'élément AC soit corrélée avec l'expression des gènes impliqués dans la biosynthèse des unités G de lignine (Raes *et al.*, 2003). Le motif PAL-box présent dans le promoteur du gène *prxC2* codant pour une peroxydase ne correspondrait alors pas exactement à un élément AC. En effet, sur les 12 paires de bases de la séquence consensus *PAL-box* reconnue par la protéine NtLIM1, les six dernières paires de base ne sont pas rigides (Kawaoka *et al.*, 2000).

## Discussion Générale

Enfin, des différences subtiles en acides aminés au sein des deux domaines LIM fixant l'ADN peuvent modifier la spécificité de fixation de la protéine sur sa séquence d'ADN cible. Ainsi, les protéines LIM au sein de chaque sous-groupe phylogénétique pourraient réguler différents gènes qui ne sont pas nécessairement impliqués dans la lignification. Par ailleurs, au sein des protéines à domaines LIM, les régions C-terminales sont très variables dans leurs compositions en acides aminés. Si le domaine activateur de la transcription est effectivement situé dans la région en C-terminal (Kawaoka *et al.*, 2000), il est alors envisageable de proposer d'autres fonctions régulatrices pour ce domaine. Ainsi, certaines protéines LIM telles que les protéines PLIM1, PLIM2,  $\delta$ LIM2 et même WLIM2, possédant une région en C-terminal très différente de celle des protéines WLIM1, pourraient à l'inverse posséder un domaine en C-terminal inhibiteur de la transcription. Pour terminer, ces facteurs ou cofacteurs de transcription LIM de plante font partie d'un complexe de transcription, et il est vraisemblable qu'ils interagissent avec d'autres protéines ou des facteurs de transcription comme ceux de type MYB. Ainsi, la régulation transcriptionnelle par les protéines à domaine LIM doit être en réalité très complexe et il est de plus probable qu'elle varie en fonction du contexte cellulaire.

## Conclusions et perspectives

### 1 Conclusions

Mon travail de thèse au sein de l'équipe « Xylème » de l'unité AGPF de l'INRA d'Orléans avait pour objectif initial d'identifier des protéines régulant la transcription de gènes impliqués dans la formation du bois de tension dont la principale caractéristique est la présence de fibres gélatineuses. Les résultats de séquençage de banques d'ADNc issues de bois tendu et opposé ont permis l'identification de nombreux gènes exprimés de façon différentielle entre ces deux tissus très contrastés (Déjardin *et al.*, 2004). L'identification de co-facteurs ou facteurs de transcription impliqués dans la formation du bois de tension permettra peut-être par la suite de contrôler l'expression d'un ensemble de gènes par la manipulation génétique d'un seul gène. En analysant la distribution des ESTs codants pour des facteurs de transcription putatifs dans les différentes banques d'ESTs de l'INRA d'Orléans, nous avons choisi d'attaquer l'étude des gènes codant pour des protéines à domaine LIM. En effet, l'expression d'un gène *LIM* (*PtaGLIM1a*) semblait être induite dans le bois de tension et ce gène est homologue au gène *NiLIM1* de tabac codant pour un facteur de transcription régulant probablement l'expression de gènes impliqués dans la voie de biosynthèse des lignines (Kawaoka *et al.*, 2000). Toutefois, des analyses bioinformatiques plus poussées ont révélé la présence de six transcrits différents codant pour des protéines possédant deux domaines LIM.

#### 1.1. Les gènes LIM de peuplier, d'*Arabidopsis* et de riz

La disponibilité de la séquence du génome de peuplier a permis de grandes avancées dans la génomique des plantes eudicotylédones. En effet, nous pouvons maintenant comparer les structures des deux génomes d'*Arabidopsis thaliana* et de *Populus trichocarpa*, et l'évolution des familles de gènes entre une plante herbacée annuelle et une plante ligneuse pérenne. De plus, les analyses de génomique comparative peuvent être étendues aux monocotylédones et mousses grâce aux génomes séquencés d'*Oryza sativa* et de *Physcomitrella patens*. Concernant la famille de gènes codant pour des protéines à domaine LIM de plante, le génome du peuplier possède douze modèles de gène *LIM* certainement tous

dupliqués. Ces six paires de gène *PtLIM* sont probablement apparues au cours de l'évolution après un événement de duplication récent. Chez les autres plantes, les familles de gènes *LIM* sont moins larges. En effet, six modèles de gène *LIM* ont été identifiés dans le génome d'*Arabidopsis*, ainsi que dans celui de riz. La mousse ne contiendrait, quant à elle, que deux modèles de gène *LIM*. De plus, l'annotation et la structure intron/exon de la séquence génomique de *Populus trichocarpa* ont été actualisées et affinées. Les gènes *LIM* de plante ont quasiment tous une structure intron/exon très conservée avec quatre introns situés à des positions bien précises dans les régions codantes de ces gènes. Cette forte conservation de la structure génomique ainsi que la forte similitude de séquence entre ces protéines au niveau des domaines LIM indiquent l'importance de ces protéines dans le fonctionnement de la cellule et la physiologie de la plante.

La famille LIM chez le tournesol, *Arabidopsis* et le tabac avait été divisée en quatre groupes phylogénétiques : PLIM1 et PLIM2 spécifiques du pollen et WLIM1 et WLIM2 exprimés dans toute la plante (Eliasson *et al.*, 2000). Or, certains gènes *LIM* découverts chez le peuplier semblent former de nouveaux groupes phylogénétiques ayant des spécificités d'expression particulières, notamment dans le bois de tension. Ainsi, afin d'améliorer la phylogénie de ces protéines à domaine LIM, j'ai décidé de collecter plus de séquences (taxons) LIM chez différentes plantes. En complément des quatre séquences génomiques de plantes actuellement disponibles, la communauté de biologistes moléculaire a misé sur l'effort de séquençage d'une quantité impressionnante d'ESTs, avec plus de 12 millions de séquences exprimées chez les plantes en Avril 2007. Ces informations de séquences ESTs m'ont permis de collecter 155 unigènes supplémentaires codant pour des protéines à domaine LIM entières chez 62 espèces représentatives des différentes familles de plantes. L'analyse phylogénétique qui en découle nous montre que les protéines à domaine LIM de plante se sont diversifiées par duplications et spéciations à partir d'un gène ancestral provenant vraisemblablement de mousses.

### *1.2. Relations entre les spécificités d'expression des gènes LIM de peuplier et la phylogénie de la famille de protéines à domaine LIM chez les plantes*

Chez les plantes vasculaires, ces protéines à domaine LIM peuvent être classées en quatre groupes  $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2, et  $\delta$ LIM2, chacun contenant des sous groupes spécifiques de familles de plantes et/ou ayant des spécificités d'expression tissulaire. Les analyses d'expression des gènes *PtaLIM* par RT-PCR semi quantitative chez le peuplier,

montrent que le profil d'expression des gènes est très diversifié, même entre gènes dupliqués. Les différences dans l'expression des gènes dupliqués suggèrent des spécificités de fonction tissulaire mais il existerait aussi de possibles redondances fonctionnelles notamment dans le pollen. Cette étude révèle un nouveau groupe de séquences apparentées appelé  $\beta$ LIM1, dont la fonction des gènes reste à explorer. Toutefois, chez les plantes, les gènes  $\beta$ LIM1 semblent être plus fortement exprimés dans les tissus indifférenciés. Chez le peuplier, les gènes dupliqués *Pt $\beta$ LIM1a* et *b* ont des profils d'expression différents, avec le gène *Pt $\beta$ LIM1a* exprimé dans une grande variété de tissus et le gène *Pt $\beta$ LIM1b* spécifiquement exprimé dans les organes reproducteurs. Au sein des sous-groupes WLIM1 et WLIM2, le peuplier possède au moins un gène exprimé de manière constitutive : les gènes *PtWLIM1a* et *PtWLIM2a*. Par contre les autres gènes dupliqués *PtWLIM1b* et *PtWLIM2b* ont respectivement une expression préférentielle dans les tissus vasculaires, et les tissus à croissance primaire incluant le bois et les feuilles. Ces analyses d'expression montrent pour la première fois une probable implication de certains gènes LIM dans le développement vasculaire de la plante. En effet, les gènes *PtGLIM1a* et *b* dupliqués sont préférentiellement exprimés dans le xylème secondaire des tiges et racines de peuplier. Ces deux gènes forment un nouveau sous-groupe appelé FLIM1 comprenant aussi le gène *GhaLIM1* de cotonnier apparemment préférentiellement exprimé dans les fibres de coton formant de la paroi secondaire. De façon remarquable, l'expression des gènes *PtGLIM1a* et *b* est également fortement induite dans le bois de tension, un tissu enrichi en cellulose. Les résultats d'expression révèlent en fait une induction plus ou moins forte d'un grand nombre de gènes LIM (*PtWLIM1a* et *b*, *PtWLIM2a*, *PtPLIM2a* et *b*, et *Pt $\delta$ LIM2a* et *b*) à la face supérieure des tiges de peupliers inclinées au niveau de différents tissus comme l'écorce, le phloème, la zone cambiale, le xylème en développement et le xylème mature. Ces profils d'expression suggèrent une fonction de ces gènes dans la perception ou la réponse de la plante au stimulus gravitationnel et/ou aux contraintes mécaniques. D'autre part, j'ai défini quatre sous groupes au sein du groupe  $\delta$ LIM2 : les sous groupes PLIM2 des eudicotylédones, rosidées et astéridées, et un nouveau sous groupe  $\delta$ LIM2 des astéridées. Les analyses d'expression confirment les analyses phylogénétiques prédisant une expression préférentielle dans le pollen des gènes *PtPLIM2a* et *b* de peuplier. Enfin, de la même manière que les gènes *PLIM2* de plante, les gènes *Pt $\delta$ LIM2a* et *b*, phylogénétiquement proches du nouveau sous groupe  $\delta$ LIM2 des astéridées, sont fortement et préférentiellement exprimés dans le pollen. Ainsi, les gènes du sous-groupe  $\delta$ LIM2 ou « PLIM2-like » des astéridées seraient peut-être aussi fortement exprimés dans le pollen. De plus la forte expression des gènes *PtPLIM2a* et *b* et *Pt $\delta$ LIM2a* et *b* dans les grains de pollen en

germination suggère une fonction de ces gènes dans la croissance du tube pollinique. De façon intéressante, ces quatre gènes et les gènes *PtWLIM1a* et *Pt $\beta$ LIM1a* sont aussi fortement exprimés dans les fibres de coton de peuplier montrant une probable implication dans la croissance de ces fibres à fort contenu en cellulose.

Par ailleurs, les protéines à domaine LIM sont également groupées en fonction de l'appartenance à une classe ou sous-classe de plante. En effet, le sous-groupe PLIM1 est spécifique de la sous-classe des astéridées, et les protéines de monocotylédones et d'eudicotylédones sont phylogénétiquement séparées pour les sous-groupes WLIM1, WLIM2 et PLIM2. De plus, bien que le groupe  $\beta$ LIM1 ne semble pas être spécifique d'une classe ou famille de plante, il ne semble pas contenir de protéines provenant d'espèces de la famille de *Brassicaceae*, comme *Arabidopsis*, et peu de graminées y sont représentées. Enfin, la diversité phylogénétique de cette famille souligne la contribution des gènes *LIM* à l'adaptation de la plante à son milieu. En effet, la spécificité d'expression pollinique des gènes appartenant aux sous-groupes PLIM1 et PLIM2, et la probable expression « fibre » spécifique des gènes du sous-groupe FLIM1 illustrent remarquablement la capacité de la plante à répondre à différentes contraintes tels que la pression de sélection gamétophytique ou le stress mécanique/gravitationnel.

### 1.3. La protéine *PtaGLIM1a* est fortement exprimée dans le bois de tension

Afin de poursuivre la caractérisation fonctionnelle de la protéine *PtGLIM1a*, nous avons décidé de la produire dans la bactérie *E. coli* sous forme de protéine recombinante, puis de la purifier. Cette protéine purifiée formerait probablement des homodimères de façon spontanée *in vitro*. Des anticorps dirigés soit contre la protéine recombinante 6His-*PtaGLIM1a*, soit contre un peptide de synthèse P-*GLIM1a* situé dans la partie C-terminale variable de la protéine ont été ensuite produits et utilisés en Western blot. L'utilisation des anticorps anti-P-*GLIM1a* confirme l'expression préférentielle de la protéine *PtaGLIM1a* dans les tissus vasculaires de peuplier et l'induction de son expression dans le bois de tension en formation ou mature. De plus, cette protéine pourrait former des homo ou hétérodimères de protéines *in vivo* et subir des modifications post-traductionnelles ressemblant à de la glycosylation. De façon intéressante, le degré de modifications post-traductionnelles semble varier en fonction de la différenciation du xylème et également en fonction de la mise en tension des tiges.

Par ce travail, j'ai pu caractériser les relations phylogénétiques entre les différents membres de la famille de protéine à domaine LIM chez les plantes et ainsi proposer une nouvelle nomenclature. L'expression des gènes *LIM* de peuplier dans les différents tissus de l'arbre nous montre des spécificités d'expression intéressantes. Cette étude a ouvert un grand nombre de champs de recherche sur les différentes fonctions putatives des gènes *LIM* de plantes. Beaucoup d'hypothèses soulevées méritent d'être approfondie, notamment celles concernant leur possible implication dans l'élongation cellulaire des tubes polliniques, des fibres vasculaires et des fibres de coton. De plus, leurs fonctions sub-cellulaires comme facteur de transcription et/ou protéine se liant aux filaments d'actine devront à l'avenir être caractérisées. Bien que les perspectives d'études sur les différents gènes *LIM* de plante soient passionnantes en regard des spécificités d'expression et des particularités structurales de ces protéines à domaine LIM, elles sont beaucoup trop nombreuses pour être exposées ici. Ainsi, je n'ouvrirai des pistes de recherches que sur le gène *PtaGLIM1a* actuellement en cours de caractérisation au laboratoire, et dont le profil d'expression suggère une implication dans la formation de la couche gélatineuse des fibres G du bois de tension.

## 2 Perspectives

### 2.1. Localisation tissulaire et cellulaire de la protéine *PtaGLIM1a*

Le gène et la protéine *PtaGLIM1a* sont fortement exprimés dans le xylème en développement et leur expression est induite dans le bois de tension chez le peuplier. Toutefois, les analyses d'expression effectuées manquent de précision de par la technique de prélèvement utilisée. Il est en effet impossible de distinguer de cette manière dans quel type cellulaire (fibres, vaisseaux ou rayons) est localisée la protéine *PtaGLIM1a*. La détermination de la spécificité d'expression cellulaire est importante afin de vérifier l'hypothèse selon laquelle *PtaGLIM1a* interviendrait lors de la différenciation des fibres gélatineuses et l'élaboration de cette couche G particulière. De plus, déterminer précisément à quel stade de différenciation du xylème la protéine *PtaGLIM1a* est exprimée est également nécessaire afin de connaître l'implication de cette protéine dans l'élongation cellulaire et/ou la formation de la paroi secondaire. C'est principalement dans cet objectif que nous avons décidé de produire une protéine recombinante *PtaGLIM1a* et d'obtenir des anticorps dirigés contre elle. Dans le même but, des anticorps dirigés contre le peptide de synthèse P-*GLIM1a* ont été produits et

permettront de détecter plus spécifiquement la protéine PtaGLIM1a notamment par rapport à son proche homologue PtaGLIM1b. Ces anticorps anti-P-GLIM1a reconnaissent bien la protéine recombinante 6His-PtaGLIM1a exprimée dans la bactérie et la protéine PtaGLIM1a native chez le peuplier. Après purification de ces anticorps par affinité contre le peptide de synthèse P-GLIM1a ou contre la protéine recombinante purifiée, ceux-ci pourront donc être utilisés pour des expériences d'immunohistlocalisation. Toutefois des difficultés ont été rencontrées lors de la purification de ces anticorps anti-P-GLIM1a contre le peptide de synthèse P-GLIM1a, et des essais de purification contre la protéine recombinante 6His-PtaGLIM1a sont en cours de réalisation au laboratoire. Ces expériences d'immunohistlocalisation permettront donc dans un premier temps d'analyser la spécificité d'expression cellulaire de la protéine PtaGLIM1a dans le xylème en formation des tiges, mais aussi plus généralement son implication éventuelle dans le système vasculaire de la plante, au niveau des nervures et pétioles de feuilles. La technique d'immunohistlocalisation est bien maîtrisée au laboratoire et a déjà permis de détecter la protéine PopFLA1 dans les fibres de bois tendu (Takeuchi *et al.*, 2007). Pour conforter ces analyses immunologiques, l'activité du promoteur du gène *PtaGLIM1a* fusionné aux gènes rapporteurs *GUS* ou *GFP* (construction promGLIM1a::*GUS* ou promGLIM1a::*GFP*) peut-être étudié dans le bois de tension de peupliers transgéniques. Des expériences similaires utilisant le promoteur de la cellulose synthase fusionné au gène rapporteur *GUS* ont déjà été réalisées au laboratoire et ont permis d'observer une forte coloration *GUS* dans le bois de tension (Déjardin *et al.* résultats non publiés). De plus, des expériences préliminaires de localisation de la *GFP* chez des peupliers génétiquement transformés avec une construction 35S::*GFP* sont prometteuses (Laurans, Communication personnelle). Enfin, en alternative, des expériences d'hybridation *in situ* peuvent être prévues pour analyser précisément la localisation du transcrit *PtaGLIM1a* dans les différents tissus de peuplier.

### 2.2. Détermination de la localisation subcellulaire de la protéine PtaGLIM1a

La détermination de la localisation subcellulaire de la protéine PtaGLIM1a est cruciale. En effet, elle permettra d'appréhender la fonction de cette protéine dans le noyau comme facteur de transcription et/ou dans le cytoplasme comme protéine associée ou non aux filaments d'actine. De plus, cette localisation subcellulaire pourrait varier en fonction de l'état de différenciation des cellules du xylème, ou de l'application d'un stimulus gravitationnel ou



mécanique. Ainsi les anticorps anti-PtaGLIM1a disponibles permettront également d'analyser par immunocytolocalisation la localisation de la protéine dans la cellule. Pour confirmer ces expériences, il sera nécessaire de produire des peupliers transgéniques portant la construction GFP-PtaGLIM1a. Enfin, au cours de ma thèse, j'ai pu mettre au point une technique de purification de noyaux basée sur un gradient de densité de percoll pour séparer les noyaux des autres organites cellulaires. Cette méthode d'isolation de noyaux nous donne une pureté en noyaux suffisante et un rendement final correct, avec respectivement  $2 * 10^8$  et  $2,3 * 10^9$  noyaux/ml pour les cellules de cals et de cambium. Ainsi, nous pourrons par la suite comparer en Western blot la quantité de protéine PtaGLIM1a dans des extraits protéiques enrichis en protéines nucléaires ou cytoplasmiques. Toutefois, cette technique de purification de noyaux n'est pas applicable sur des cellules de xylème possédant une paroi lignifiée, car cette dernière co-sédimente avec la fraction de noyaux.

### *2.3. La protéine PtaGLIM1a est elle un facteur de transcription et/ou une protéine se liant à l'actine ?*

La protéine NtWLIM1 de tabac est capable de se fixer à l'ADN et d'activer l'expression de gènes de la voie de biosynthèse des lignines (Kawaoka *et al.*, 2000). C'est aussi pour analyser la fixation de la protéine PtaGLIM1a à l'ADN par des expériences de gel retard (EMSA) que nous avons décidé de produire la protéine recombinante 6His-PtaGLIM1a et de la purifier. Nous pourrons dans un premier temps, utiliser une sonde oligonucléotide correspondant au motif *PAL-box* reconnu par la protéine NtWLIM1 (Kawaoka *et al.*, 2000), ou bien grâce à la séquence du génome de peuplier, rechercher dans les promoteurs des gènes de la voie de biosynthèse des lignines un motif consensus correspondant à l'élément AC. De plus, afin de rechercher le motif consensus de fixation à l'ADN par la protéine PtaGLIM1a sans avoir des a priori sur la séquence cible, des expériences de « sélection par PCR du site de fixation » pourront être menées en utilisant une sonde oligonucléotide dégénérée (Johannesson *et al.*, 2001). De même, les sites ADN de fixation de la protéine PtaGLIM1a pourront être analysés en utilisant la méthode « whole genome PCR » (Kinzler & Vogelstein, 1989). Par ailleurs, il sera possible de réaliser des expériences de « super-shift » avec les anticorps anti-PtaGLIM1a. Enfin les anticorps seront nécessaires pour isoler le facteur de transcription lors des expériences d'immunoprécipitation de chromatine (ChIP) actuellement en développement au laboratoire. Cette technique permet par la fixation covalente (pontage) des protéines à l'ADN *in vivo*, puis par la sonication de la chromatine en des fragments de

petite taille, d'isoler les séquences cibles du facteur ou cofacteur de transcription grâce à des anticorps spécifiquement dirigés contre celui-ci. Pour cette technique, j'ai pu mettre au point la première étape de purification de noyaux. De plus, les expériences de sonication de la chromatine sur du matériel non ponté génèrent des fragments d'ADN de bonne taille entre 500 et 1000 pb. Toutefois, le pontage de la chromatine par le formaldéhyde est difficile sur les cellules de cambium et de xylème. En effet, j'observe après pontage une forte dégradation de l'ADN générant des fragments de différentes tailles. Cet effet peut-être dû à une interaction entre le formaldéhyde et les phénols du bois entraînant une oxydation du matériel.

Il a été montré que la protéine NtWLIM1 se lie aux filaments d'actine et les empaquette (Thomas *et al.*, 2006). La détermination de la localisation cytoplasmique de la protéine PtaGLIM1a par des expériences d'immunocytolocalisation et de colocalisation avec l'actine et la tubuline, nous indiquera son association ou non avec ces deux éléments du cytosquelette. De plus, la latrunculine B et l'oryzaline affectant respectivement les réseaux d'actine et de microtubules pourront être utilisées pour estimer leur effet sur la localisation de la protéine PtaGLIM1a. En alternative, nous pourrions aussi co-transformer des peupliers avec à la fois la protéine PtaGLIM1a fusionnée à la GFP et l'actine fusionnée à la RFP ou la YFP, pour regarder leur éventuelle colocalisation dans les cellules. Avec ce double mutant, il serait également envisageable d'utiliser la technique de transfert d'énergie entre molécules fluorescentes (FRET) pour analyser l'interaction entre la protéine PtaGLIM1a et l'actine. Ces expériences *in vivo* sont peut-être difficilement applicables sur du tissu bois, et l'utilisation de système de culture cellulaire de peuplier sera peut être plus adaptée. De plus, afin de déterminer les propriétés de fixation de la protéine PtaGLIM1a à l'actine *in vitro*, des expériences de cosédimentation à haute vitesse avec l'actine seront nécessaires. Enfin, l'interaction de la protéine PtaGLIM1a avec d'autres protéines est aussi probable. Pour déterminer ces partenaires dans la cellule, le système « double hybride » chez la levure peut-être testé pour récupérer les protéines se liant à PtaGLIM1a. Enfin, des expériences de co-immunoprécipitation conforteront les interactions protéine/protéine déterminées.

### *2.4. Fonction de la protéine PtaGLIM1a dans la formation de bois de tension*

Une des manières d'étudier la fonction d'un gène dans la plante est d'abaisser (*knock-down*), de supprimer (*knock-out*), ou inversement d'augmenter (*overexpression*) le niveau d'expression du gène, puis d'examiner le phénotype des plantes transgéniques. Contrairement

## Conclusions et Perspectives

à *Arabidopsis*, des lignées *knock-out* d'insertion ADN-T ne sont pas disponibles pour le peuplier. C'est pourquoi, pour diminuer l'expression du gène *PtaGLIM1a* chez le peuplier, d'autres stratégies ont été envisagées comme la création de lignées RNAi (*RNA interference*) ou des lignées portant une construction antisens *PtaGLIM1a*. Nous avons rencontré des difficultés pour construire un vecteur de type RNAi (vecteur pHELLSGATE) pour ce gène en utilisant le système de clonage Gateway. Toutefois, des lignées de peupliers (clone 717-1-B4) transgéniques surexprimant le gène *PtaGLIM1a* sous le contrôle du promoteur double 35S (vecteur pMDC32) sont en cours d'analyse au laboratoire. De façon surprenante, certaines lignées présentent des feuilles déformées ou un phénotype « gaufré » similaire à celui observé sur les feuilles de plantes d'*Arabidopsis* surexprimant le gène *SND1* (Zhong *et al.*, 2006). Toutefois, ce phénotype n'est apparu que sur de jeunes plantules cultivées en chambre de culture, et le niveau d'expression du gène *PtaGLIM1a* n'a pas encore été analysé chez ces lignées. Nous ne savons donc pas si ce phénotype provient de la transformation génétique ou des conditions de cultures. Néanmoins, ce gène pourrait éventuellement agir sur la morphologie des cellules chez le peuplier. Lors des futures analyses de ces peupliers transgéniques, il sera indispensable de vérifier le niveau d'expression du gène et de la protéine PtaGLIM1a, et d'évaluer l'effet de la dérégulation du gène sur la structure des filaments d'actine. De plus, la surexpression de la protéine PtaGLIM1a chez le peuplier a peut être une conséquence sur la formation et le maintien de câbles de microfilaments d'actine, de la même manière que la protéine NtWLIM1 dans les cellules BY2 de tabac (Thomas *et al.*, 2006). De même, les effets sur le cytosquelette de microtubules seront regardés. Par ailleurs, des analyses microarray sur ces lignées transgéniques pourraient nous indiquer quels sont les gènes régulés par ce facteur de transcription putatif. De plus, nous examinerons bien sûr la morphologie des cellules du xylème, la structure des parois cellulaires, et la teneur en lignine et cellulose ainsi que la composition en lignine des tiges. Enfin, nous pourrions estimer l'effet de la surexpression du gène *PtaGLIM1a* sur l'induction de la formation des fibres gélatineuses de bois de tension après inclinaison des tiges.

## Conclusions et Perspectives

## Bibliographie

- Abe H., Funada R., Ohtani J. & Fukazawa K. (1995) Changes in the arrangement of microtubules and microfibrils in differentiating conifer tracheids during the expansion of cells. *Annals of Botany* **75**: 305-310
- Agreste (2004) -Ministère de l'Agriculture et de la Pêche- Production de bois 2004. *Données internet disponibles sur* : [http://agreste.agriculture.gouv.fr/thematiques\\_7/bois\\_forets\\_88/index.html](http://agreste.agriculture.gouv.fr/thematiques_7/bois_forets_88/index.html)
- Allina S. M., Pri-Hadash A., Theilmann D. A., Ellis B. E. & Douglas C. J. (1998) 4-coumarate:coenzyme A ligase in hybrid poplar. properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. *Plant Physiology* **116**: 743-754
- Allona I., Quinn M., Shoop E., Swope K., Cyr S. S., Carlis J., Riedl J., Retzel E., Campbell M. M., Sederoff R., *et al.* (1998) Analysis of xylem formation in pine by cDNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 9693-9698
- Altschul S. F., Gish W., Miller W., Myers E. W. & Lipman D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410
- Amagai M., Ariizumi T., Endo M., Hatakeyama K., Kuwata C., Shibata D., Toriyama K. & Watanabe M. (2003) Identification of anther-specific genes in a cruciferous model plant, *Arabidopsis thaliana*, by using a combination of *Arabidopsis* macroarray and mRNA derived from *Brassica oleracea*. *Sexual Plant Reproduction* **15**: 213-220
- Anderhag P., Hepler P. K. & Lazzaro M. D. (2000) Microtubules and microfilaments are both responsible for pollen tube elongation in the conifer *Picea abies* (Norway spruce). *Protoplasma* **214**: 141-157
- Andersson-Gunneras S., Mellerowicz E. J., Love J., Segerman B., Ohmiya Y., Coutinho P. M., Nilsson P., Henrissat B., Moritz T. & Sundberg B. (2006) Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *The Plant Journal* **45**: 144-165
- Anterola A. M., Jeon J.-H., Davin L. B. & Lewis N. G. (2002) Transcriptional control of monolignol biosynthesis in *Pinus taeda*. Factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *Journal of Biological Chemistry* **277**: 18272-18280
- Aoyama W., Matsumura A., Tsutsumi Y. & Nishida T. (2001) Lignification and peroxidase in tension wood of *Eucalyptus viminalis* seedlings. *Journal of Wood Science* **47**: 419-424
- Arber S. & Caroni P. (1996) Specificity of single LIM motifs in targeting and LIM/LIM interactions *in situ*. *Genes & Development* **10**: 289-300
- Arber S., Halder G. & Caroni P. (1994) Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation. *Cell* **79**: 221-231
- Arber S., Hunter J. J., Ross J., Hongo M., Sansig G., Borg J., Perriard J.-C., Chien K. R. & Caroni P. (1997) MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell* **88**: 393-403
- Arechaga I., Miroux B., Karrasch S., Huijbregts R., de Kruijff B., Runswick M. J. & Walker J. E. (2000) Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F1Fo ATP synthase. *FEBS Letters* **482**: 215-219
- Arioli T., Peng L., Betzner A. S., Burn J., Wittke W., Herth W., Camilleri C., Höfte H., Plazinski J., Birch R., *et al.* (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science* **279**: 717-720

## Bibliographie

- Arnaud D., Déjardin A., Leplé J.-C., Lesage-Descauses M.-C. & Pilate G. (2007) Genome-wide analysis of *LIM* gene family in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*. *DNA Research* **14**: 103-116
- Bach I. (2000) The LIM domain: regulation by association. *Mechanisms of Development* **91**: 5-17
- Bailey T. L. & Elkan C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / International Conference on Intelligent Systems for Molecular Biology* **2**: 28-36
- Baltz R., Domon C., Pillay D. & Steinmetz A. (1992a) Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein. *The Plant Journal* **2**: 713-721
- Baltz R., Evrard J.-L., Bourdon V. & Steinmetz A. (1996) The pollen-specific LIM protein PLIM-1 from sunflower binds nucleic acids in vitro. *Sexual Plant Reproduction* **9**: 264-268
- Baltz R., Evrard J. L., Domon C. & Steinmetz A. (1992b) A LIM motif is present in a pollen-specific protein. *Plant Cell* **4**: 1465-1466
- Baltz R., Schmit A.-C., Kohnen M., Hentges F. & Steinmetz A. (1999) Differential localization of the LIM domain protein PLIM-1 in microspores and mature pollen grains from sunflower. *Sexual Plant Reproduction* **12**: 60-65
- Baluska F., Samaj J., Wojtaszek P., Volkmann D. & Menzel D. (2003) Cytoskeleton-plasma membrane-cell wall continuum in plants. Emerging links revisited. *Plant Physiology* **133**: 482-491
- Bamber R. K. (1987) The origin of growth stresses: a rebuttal. *IAWA Bulletin* **8**: 80-84
- Bate N., Orr J., Ni W., Meromi A., Nadler-Hassar T., Doerner P., Dixon R., Lamb C. & Elkind Y. (1994) Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate-determining step in natural product synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 7608-7612
- Baucher M., Chabbert B., Pilate G., Van Doorselaere J., Tollier M. T., Petit-Conil M., Cornu D., Monties B., Van Montagu M., Inze D., *et al.* (1996) Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar. *Plant Physiology* **112**: 1479-1490
- Becker J. D., Boavida L. C., Carneiro J., Haury M. & Feijo J. A. (2003) Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiology* **133**: 713-725
- Besseau S., Hoffmann L., Geoffroy P., Lapierre C., Pollet B. & Legrand M. (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* **19**: 148-162
- Bevan M., Shufflebottom D., Edwards K., Jefferson R. & Schuch W. (1989) Tissue- and cell-specific activity of a phenylalanine ammonia-lyase promoter in transgenic plants. *EMBO Journal* **8**: 1899-1906
- Bhandari S., Fujino T., Thammanagowda S., Zhang D., Xu F. & Joshi C. (2006) Xylem-specific and tension stress-responsive coexpression of KORRIGAN endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. *Planta* **224**: 828-837
- Birney E. & Durbin R. (2000) Using GeneWise in the *Drosophila* annotation experiment. *Genome Research* **10**: 547-548
- Blancaflor E. B. & Masson P. H. (2003) Plant gravitropism. Unraveling the ups and downs of a complex process. *Plant Physiology* **133**: 1677-1690
- Boerjan W., Ralph J. & Baucher M. (2003) Lignin biosynthesis. *Annual Review of Plant Biology* **54**: 519-546

## Bibliographie

- Boes T. K. & Strauss S. H. (1994) Floral phenology and morphology of black cottonwood, *Populus trichocarpa* (Salicaceae). *American Journal of Botany* **81**: 562-567
- Bonke M., Thitamadee S., Mahonen A. P., Hauser M.-T. & Helariutta Y. (2003) APL regulates vascular tissue identity in *Arabidopsis*. *Nature* **426**: 181-186
- Boudet A. M., Kajita S., Grima-Pettenati J. & Goffner D. (2003) Lignins and lignocellulosics: a better control of synthesis for new and improved uses. *Trends in Plant Science* **8**: 576-581
- Bouquin T., Mattsson O., Naested H., Foster R. & Mundy J. (2003) The *Arabidopsis lue1* mutant defines a katanin p60 ortholog involved in hormonal control of microtubule orientation during cell growth. *Journal of Cell Science* **116**: 791-801
- Bradford M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254
- Bradshaw H. D., Ceulemans R., Davis J. & Stettler R. (2000) Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. *Journal of Plant Growth Regulation* **19**: 306-313
- Brière C., Barthou H. & Petitprez M. (2004) A new tool for plant cell biology: *in vivo* antibody uptake in plant protoplasts. *Plant Cell Reports* **22**: 878-884
- Brière C., Bordel A.-C., Barthou H., Jauneau A., Steinmetz A., Alibert G. & Petitprez M. (2003) Is the LIM-domain protein HaWLIM1 associated with cortical microtubules in sunflower protoplasts? *Plant and Cell Physiology* **44**: 1055-1063
- Brosche M., Vinocur B., Alatalo E., Lamminmaki A., Teichmann T., Ottow E., Djilianov D., Afif D., Bogeat-Triboulot M.-B., Altman A., *et al.* (2005) Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. *Genome Biology* **6**: R101
- Brown D. E., Rashotte A. M., Murphy A. S., Normanly J., Tague B. W., Peer W. A., Taiz L. & Muday G. K. (2001) Flavonoids act as negative regulators of auxin transport *in vivo* in *Arabidopsis*. *Plant Physiology* **126**: 524-535
- Brown M. R., Saxena I. M. & Kudlicka K. (1996) Cellulose biosynthesis in higher plants. *Trends in Plant Science* **1**: 149-156
- Buchberger A., Schroder H., Hesterkamp T., Schonfeld H.-J. & Bukau B. (1996) Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. *Journal of Molecular Biology* **261**: 328-333
- Burk D. H., Liu B., Zhong R., Morrison W. H. & Ye Z.-H. (2001) A katanin-like protein regulates normal cell wall biosynthesis and cell elongation. *Plant Cell* **13**: 807-828
- Burk D. H. & Ye Z.-H. (2002) Alteration of oriented deposition of cellulose microfibrils by mutation of a katanin-like microtubule-severing protein. *Plant Cell* **14**: 2145-2160
- Burlat V., Kwon M., Davin L. B. & Lewis N. G. (2001) Dirigent proteins and dirigent sites in lignifying tissues. *Phytochemistry* **57**: 883-897
- Busov V. B., Brunner A. M., Meilan R., Filichkin S., Ganio L., Gandhi S. & Strauss S. H. (2005) Genetic transformation: a powerful tool for dissection of adaptive traits in trees. *New phytologist* **167**: 9-18
- Cattaruzza M., Lattrich C. & Hecker M. (2004) Focal adhesion protein zyxin is a mechanosensitive modulator of gene expression in vascular smooth muscle cells. *Hypertension* **43**: 726-730
- Cervera M. T., Storme V., Soto A., Ivens B., Montagu M. V., Rajora O. P. & Boerjan W. (2005) Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. *Theoretical and Applied Genetics* **111**: 1440-1456

## Bibliographie

- Chaffey N. (2002) Wood formation in trees: cell and molecular biology techniques. In N Chaffey, ed, Wood formation in trees: cell and molecular biology techniques. Taylor & Francis Ltd; London; UK
- Chaffey N. & Barlow P. (2001) The cytoskeleton facilitates a three-dimensional symplasmic continuum in the long-lived ray and axial parenchyma cells of angiosperm trees. *Planta* **213**: 811-823
- Chaffey N. & Barlow P. (2002) Myosin, microtubules, and microfilaments: co-operation between cytoskeletal components during cambial cell division and secondary vascular differentiation in trees. *Planta* **214**: 526-536
- Chaffey N., Barlow P. & Barnett J. (1997a) Cortical microtubules rearrange during differentiation of vascular cambial derivatives, microfilaments do not. *Trees - Structure and Function* **11**: 333-341
- Chaffey N., Barlow P. & Barnett J. (2000a) A cytoskeletal basis for wood formation in angiosperm trees: the involvement of microfilaments. *Planta* **210**: 890-896
- Chaffey N., Barlow P. & Barnett J. (2000b) Structure-function relationships during secondary phloem development in an angiosperm tree, *Aesculus hippocastanum*: microtubules and cell walls. *Tree Physiology* **20**: 777-786
- Chaffey N., Barlow P. & Sundberg B. (2002) Understanding the role of the cytoskeleton in wood formation in angiosperm trees: hybrid aspen (*Populus tremula* × *P. tremuloides*) as the model species. *Tree Physiology* **22**: 239-249
- Chaffey N., Barnett J. & Barlow P. (1999) A cytoskeletal basis for wood formation in angiosperm trees: the involvement of cortical microtubules. *Planta* **208**: 19-30
- Chaffey N. J., Barnett J. R. & Barlow P. W. (1997b) Cortical microtubule involvement in bordered pit formation in secondary xylem vessel elements of *Aesculus hippocastanum* L. (Hippocastanaceae): A correlative study using electron microscopy and indirect immunofluorescence microscopy. *Protoplasma* **197**: 64-75
- Chang D. F., Belaguli N. S., Chang J. & Schwartz R. J. (2007) LIM-only protein, CRP2, switched on smooth muscle gene activity in adult cardiac myocytes. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 157-162
- Chang D. F., Belaguli N. S., Iyer D., Roberts W. B., Wu S.-P., Dong X.-R., Marx J. G., Moore M. S., Beckerle M. C., Majesky M. W., *et al.* (2003) Cysteine-Rich LIM-Only proteins CRP1 and CRP2 are potent smooth muscle differentiation cofactors. *Developmental Cell* **4**: 107-118
- Chang S. J., Puryear J. & Cairney J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**: 113-116
- Chen C., Meyermans H., Burggraef B., De Rycke R. M., Inoue K., De Vleeschauwer V., Steenackers M., Van Montagu M. C., Engler G. J. & Boerjan W. A. (2000) Cell-specific and conditional expression of caffeoyl-coenzyme A-3-O-methyltransferase in poplar. *Plant Physiology* **123**: 853-868
- Chen Y., Chen T., Shen S., Zheng M., Guo Y., Lin J., Baluska F. & Samaj J. (2006) Differential display proteomic analysis of *Picea meyeri* pollen germination and pollen-tube growth after inhibition of actin polymerization by latrunculin B. *The Plant Journal* **47**: 174-195
- Chenna R., Sugawara H., Koike T., Lopez R., Gibson T. J., Higgins D. G. & Thompson J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **31**: 3497-3500
- Cheung A. Y., Wang H. & Wu H.-m. (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. *Cell* **82**: 383-393
- Christensen H. E. M., Ramachandran S., Tan C.-T., Surana U., Dong C.-H. & Chua N.-H. (1996) *Arabidopsis* profilins are functionally similar to yeast profilins: identification



## Bibliographie

- of a vascular bundle-specific profilin and a pollen-specific profilin. *The Plant Journal* **10**: 269-279
- Chu Z., Chen H., Zhang Y., Zhang Z., Zheng N., Yin B., Yan H., Zhu L., Zhao X., Yuan M., *et al.* (2007) Knockout of the *AtCESA2* gene affects microtubule orientation and causes abnormal cell expansion in *Arabidopsis*. *Plant Physiology* **143**: 213-224
- Clair B., Almeras T., Yamamoto H., Okuyama T. & Sugiyama J. (2006) Mechanical behavior of cellulose microfibrils in tension wood, in relation with maturation stress generation. *Biophysical Journal* **91**: 1128-1135
- Cline J., Braman J. & Hogrefe H. (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research* **24**: 3546-3551
- Collin E. & Gilbert J. M. (1995) Comportement de peupliers grisard et d'hybrides de tremble en jeunes plantations. *Ingénieries* **1**: 3-10
- Cosgrove D. J. (1997) Assembly and enlargement of the primary cell wall in plants. *Annual Review of Cell and Developmental Biology* **13**: 171-201
- Craigon D. J., James N., Okyere J., Higgins J., Jotham J. & May S. (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Research* **32**: D575-577
- Crawford A., Pino J. & Beckerle M. (1994) Biochemical and molecular characterization of the chicken cysteine-rich protein, a developmentally regulated LIM-domain protein that is associated with the actin cytoskeleton. *Journal of Cell Biology* **124**: 117-127
- Cvrcková F., Rivero F. & Bavlínka B. (2004) Evolutionarily conserved modules in actin nucleation: lessons from *Dictyostelium discoideum* and plants. *Protoplasma* **224**: 15-31
- Damerval C., De Vienne D., Zivy M. & Thiellement H. (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis* **7**: 52-54
- Dawid I. B., Breen J. J. & Toyama R. (1998) LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends in Genetics* **14**: 156-162
- Dean Rider S., Henderson J. T., Jerome R. E., Edenberg H. J., Romero-Severson J. & Ogas J. (2003) Coordinate repression of regulators of embryonic identity by *PICKLE* during germination in *Arabidopsis*. *The Plant Journal* **35**: 33-43
- Déjardin A., Leplé J.-C., Lesage-Descauses M.-C., Costa G. & Pilate G. (2004) Expressed sequence tags from poplar wood tissues - a comparative analysis from multiple libraries. *Plant Biology* **6**: 55-64
- Delmer D. P. & Amor Y. (1995) Cellulose biosynthesis. *Plant Cell* **7**: 987-1000
- Demura T. & Fukuda H. (2007) Transcriptional regulation in wood formation. *Trends in Plant Science* **12**: 64-70
- Djerbi S., Aspeborg H., Nilsson P., Sundberg B., Mellerowicz E., Blomqvist K. & Teeri T. T. (2004) Identification and expression analysis of genes encoding putative cellulose synthases (CesA) in the hybrid aspen, *Populus tremula* (L.) × *P. tremuloides* (Michx.). *Cellulose* **11**: 301-312
- Djerbi S., Lindskog M., Arvestad L., Sterky F. & Teeri T. (2005) The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (CesA) genes. *Planta* **221**: 739-746
- Donaldson L. A. (2001) Lignification and lignin topochemistry -- an ultrastructural view. *Phytochemistry* **57**: 859-873
- Doolittle R. F. (1992) Reconstructing history with amino acid sequences. *Protein Science* **1**: 191-200
- Douglas C. J. (1996) Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. *Trends in Plant Science* **1**: 171-178

## Bibliographie

- Du J., Xie H.-L., Zhang D.-Q., He X.-Q., Wang M.-J., Li Y.-Z., Cui K.-M. & Lu M.-Z. (2006) Regeneration of the secondary vascular system in poplar as a novel system to investigate gene expression by a proteomic approach. *PROTEOMICS* **6**: 881-895
- Dvornyk V., Sirvio A., Mikkonen M. & Savolainen O. (2002) Low nucleotide diversity at the *pall* locus in the widely distributed *Pinus sylvestris*. *Molecular Biology and Evolution* **19**: 179-188
- Eckenwalder J. E. (1996) Systematics and evolution of Populus. In RF Stettler, Bradshaw Jr., H.D., Heilman, P.E., Hinckley, T.M., ed, Biology of Populus and its implication for management and conservation. NRC Research Press, Ottawa, 7-32
- Ehltng J., Buttner D., Wang Q., Douglas C. J., Somssich I. E. & Kombrink E. (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *The Plant Journal* **19**: 9-20
- Eliasson Å., Gass N., Mundel C., Baltz R., Kräuter R., Evrard J.-L. & Steinmetz A. (2000) Molecular and expression analysis of a LIM protein gene family from flowering plants. *Molecular and General Genetics* **264**: 257-267
- Elkind Y., Edwards R., Mavandad M., Hedrick S., Ribak O., Dixon R. & Lamb C. (1990) Abnormal plant development and down-regulation of phenylpropanoid biosynthesis in transgenic tobacco containing a heterologous phenylalanine ammonia-lyase gene. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 9057-9061
- Ellis R. J. (2005) Chaperomics: *In vivo* GroEL function defined. *Current Biology* **15**: R661-R663
- Endo M., Tsuchiya T., Saito H., Matsubara H., Hakozaiki H., Masuko H., Kamada M., Higashitani A., Takahashi H., Fukuda H., *et al.* (2004) Identification and molecular characterization of novel anther-specific genes in *Oryza sativa* L. by using cDNA microarray. *Genes & Genetic Systems* **79**: 213-226
- Eudes A., Pollet B., Sibout R., Do C.-T., Séguin A., Lapierre C. & Jouanin L. (2006) Evidence for a role of *AtCAD 1* in lignification of elongating stems of *Arabidopsis thaliana*. *Planta* **225**: 23-39
- Feuerstein R., Wang X., Song D., Cooke N. & Liebhaber S. (1994) The LIM/double zinc-finger motif functions as a protein dimerization domain. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 10655-10659
- Flick M. & Konieczny S. (2000) The muscle regulatory and structural protein MLP is a cytoskeletal binding partner of beta-spectrin. *Journal of Cell Science* **113**: 1553-1564
- Fournier M., Baillères H. & Chanson B. (1994) Tree biomechanics: growth, cumulative prestresses, and reorientations. *Biomimetics* **2**: 229-251
- Freyd G., Kim S. K. & Horvitz H. R. (1990) Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-II*. *Nature* **344**: 876-879
- Fukuda H. (1996) Xylogenesis: initiation, progression, and cell death. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 299-325
- Funada R., Abe H., Furusawa O., Imaizumi H., Fukazawa K. & Ohtani J. (1997) The orientation and localization of cortical microtubules in differentiating conifer tracheids during cell expansion. *Plant Cell Physiology* **38**: 210-212
- Gamborg O. L., Miller R. A. & Ojima K. (1968) Nutrient requirements of suspension cultures of soybean root cells. *Experimental Cell Research* **50**: 151-158
- Gardiner J. C., Taylor N. G. & Turner S. R. (2003) Control of cellulose synthase complex localization in developing xylem. *Plant Cell* **15**: 1740-1748
- Gascuel O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* **14**: 685-695

## Bibliographie

- Gatti S., Shen W. H., Grausem B., Dieterle M., Hoffmann C., Thomas C. & Steinmetz A. (2005) Subcellular localisation studies of the tobacco LIM domain protein NtWLM2. *FEBS/ESF Workshop on Integrated Approaches in Cytoskeleton Research, Luxembourg*
- Geitmann A. & Parre E. (2004) The local cytomechanical properties of growing pollen tubes correspond to the axial distribution of structural cellular elements. *Sexual Plant Reproduction* **17**: 9-16
- Gepstein S., Sabeji G., Carp M.-J., Hajouj T., Nesher M. F. O., Yariv I., Dor C. & Bassani M. (2003) Large-scale identification of leaf senescence-associated genes. *The Plant Journal* **36**: 629-642
- Gibbon B. C., Kovar D. R. & Staiger C. J. (1999) Latrunculin B has different effects on pollen germination and tube growth. *Plant Cell* **11**: 2349-2364
- Gierlinger N. & Schwanninger M. (2006) Chemical imaging of poplar wood cell walls by confocal Raman microscopy. *Plant Physiology* **140**: 1246-1254
- Gill G. & Ptashne M. (1987) Mutants of GAL4 protein altered in an activation function. *Cell* **51**: 121-126
- Gion J.-M., Lalanne C., Le Provost G., Ferry-Dumazet H., Paiva J., Chaumeil P., Frigerio J.-M., Brach J., Barré A., de Daruvar A., *et al.* (2005) The proteome of maritime pine wood forming tissue. *PROTEOMICS* **5**: 3731-3751
- Gomez-Maldonado J., Avila C., Torre F., Canas R., Canovas F. M. & Campbell M. M. (2004) Functional interactions between a glutamine synthetase promoter and MYB proteins. *The Plant Journal* **39**: 513-526
- Gorovoy M., Niu J., Bernard O., Profirovic J., Minshall R., Neamu R. & Voyno-Yasenetskaya T. (2005) LIM kinase 1 coordinates microtubule stability and actin polymerization in human endothelial cells. *Journal of Biological Chemistry* **280**: 26533-26542
- Gossot O. & Geitmann A. (2007) Pollen tube growth: coping with mechanical obstacles involves the cytoskeleton. *Planta* **226**: 405-416
- Gournier H., Goley E. D., Niederstrasser H., Trinh T. & Welch M. D. (2001) Reconstitution of Human Arp2/3 complex reveals critical roles of individual subunits in complex structure and activity. *Molecular Cell* **8**: 1041-1052
- Gray-Mitsumune M., Molitor E., Cukovic D., Carlson J. & Douglas C. (1999) Developmentally regulated patterns of expression directed by poplar *PAL* promoters in transgenic tobacco and poplar. *Plant Molecular Biology* **39**: 657-669
- Groot T. T., Bodegom P. M. v., Meijer H. A. J. & Harren F. J. M. (2005) Gas transport through the root-shoot transition zone of rice tillers. *Plant and Soil* **277**: 107-116
- Groover A., Mansfield S., DiFazio S., Dupper G., Fontana J., Millar R. & Wang Y. (2006) The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Molecular Biology* **61**: 917-932
- Groover A. T. (2005) What genes make a tree a tree? *Trends in Plant Science* **10**: 210-214
- Grubinger M. & Gimona M. (2004) CRP2 is an autonomous actin-binding protein. *FEBS Letters* **557**: 88-92
- Grünwald C., Ruel K., Kim Y. S. & Schmitt U. (2002) On the cytochemistry of cell wall formation in poplar trees. *Plant Biology* **4**: 13-21
- Guillaumie S., Pichon M., Martinant J.-P., Bosio M., Goffner D. & Barrière Y. (2007) Differential expression of phenylpropanoid and related genes in brown-midrib *bm1*, *bm2*, *bm3*, and *bm4* young near-isogenic maize plants. *Planta* **226**: 235-250
- Guindon S. & Gascuel O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696 - 704

## Bibliographie

- Guyon V. N., Astwood J. D., Garner E. C., Dunker A. K. & Taylor L. P. (2000) Isolation and characterization of cDNAs expressed in the early stages of flavonol-induced pollen germination in *Petunia*. *Plant Physiology* **123**: 699-710
- Hamberger B. & Hahlbrock K. (2004) The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 2209-2214
- Hamzeh M. & Dayanandan S. (2004) Phylogeny of *Populus* (*Salicaceae*) based on nucleotide sequences of chloroplast *TRNT-TRNF* region and nuclear rDNA. *American Journal of Botany* **91**: 1398-1408
- Harada J. J. (2001) Role of *Arabidopsis* *LEAFY COTYLEDON* genes in seed development. *Journal of Plant Physiology* **158**: 405-409
- Harding S. A., Leshkevich J., Chiang V. L. & Tsai C.-J. (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiology* **128**: 428-438
- Harper B. D., Beckerle M. C. & Pomiès P. (2000) Fine mapping of the alpha-actinin binding site within cysteine-rich protein. *Biochemical Journal* **350**: 269-274
- Hatton D., Sablowski R., Yung M.-H., Smith C., Schuch W. & Bevan M. (1995) Two classes of *cis* sequences contribute to tissue-specific expression of a *PAL2* promoter in transgenic tobacco. *The Plant Journal* **7**: 859-876
- Hellgren J. M., Olofsson K. & Sundberg B. (2004) Patterns of auxin distribution during gravitational induction of reaction wood in poplar and pine. *Plant Physiology* **135**: 212-220
- Hepler P. K., Vidali L. & Cheung A. Y. (2001) Polarized cell growth in higher plants. *Annual Review of Cell and Developmental Biology* **17**: 159-187
- Herreros L., Rodriguez-Fernandez J. L., Brown M. C., Alonso-Lebrero J. L., Cabanas C., Sanchez-Madrid F., Longo N., Turner C. E. & Sanchez-Mateos P. (2000) Paxillin localizes to the lymphocyte microtubule organizing center and associates with the microtubule cytoskeleton. *Journal of Biological Chemistry* **275**: 26436-26440
- Hertzberg M., Aspeborg H., Schrader J., Andersson A., Erlandsson R., Blomqvist K., Bhalerao R., Uhlen M., Teeri T. T., Lundeberg J., *et al.* (2001) A transcriptional roadmap to wood formation. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 14732-14737
- Higuchi T., Ito T., Umezawa T., Hibino T. & Shibata D. (1994) Red-brown color of lignified tissues of transgenic plants with antisense CAD gene: Wine-red lignin from coniferyl aldehyde. *Journal of Biotechnology* **37**: 151-158
- Hoffmann C., Thomas C. & Steinmetz A. (2005) The tobacco LIM protein NtWLM1 stabilises the actin cytoskeleton against depolymerisation by latrunculin B. *FEBS/ESF Workshop on Integrated Approaches in Cytoskeleton Research, Luxembourg*
- Honys D. & Twell D. (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiology* **132**: 640-652
- Honys D. & Twell D. (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. *Genome Biology* **5**: R85
- Hu W.-J., Harding S. A., Lung J., Popko J. L., Ralph J., Stokke D. D., Tsai C.-J. & Chiang V. L. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology* **17**: 808-812
- Hu W.-J., Kawaoka A., Tsai C.-J., Lung J., Osakabe K., Ebinuma H. & Chiang V. L. (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). *Proceedings of the National Academy of Sciences of the United States of America* **95**: 5407-5412

## Bibliographie

- Huang C. Z., Wu X. L., Jiang C. Z., Hu Z. H., He H. Y., Liou Z. H., Xu Y. H., Jin W., Zheng T. & Chen J. Q. (2007) RETRACTED: Isolation, characterization and expression of a gene from *Brassica napus* encoding a LIM-domain protein. *Plant Physiology and Biochemistry*: in press
- Huang S., McDowell J. M., Weise M. J. & Meagher R. B. (1996) The *Arabidopsis* profilin gene family (evidence for an ancient split between constitutive and pollen-specific profilin genes). *Plant Physiology* **111**: 115-126
- Huang S., Robinson R. C., Gao L. Y., Matsumoto T., Brunet A., Blanchoin L. & Staiger C. J. (2005) Arabidopsis VILLIN1 generates actin filament cables that are resistant to depolymerization. *Plant Cell* **17**: 486-501
- Ingvarsson P. K. (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., *Salicaceae*). *Genetics* **169**: 945-953
- Isebrands J. G. & Bense D. W. (1972) Incidence and structure of gelatinous fibers within rapid-growing Eastern Cottonwood. *Wood and Fiber* **4**: 61-71
- Jaligot E., Beulé T. & Rival A. (2002) Methylation-sensitive RFLPs: characterisation of two oil palm markers showing somaclonal variation-associated polymorphism. *Theoretical and Applied Genetics* **104**: 1263-1269
- Jansson S. & Douglas C. J. (2007) Populus: A model system for plant biology. *Annual Review of Plant Biology* **58**: 435-458
- Johannesson H., Wang Y. & Engström P. (2001) DNA-binding and dimerization preferences of *Arabidopsis* homeodomain-leucine zipper transcription factors *in vitro*. *Plant Molecular Biology* **45**: 63-73
- Jones D. T., Taylor W. R. & Thornton J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**: 275-282
- Jones M. A., Raymond M. J. & Smirnov N. (2006) Analysis of the root-hair morphogenesis transcriptome reveals the molecular identity of six genes with roles in root-hair development in *Arabidopsis*. *The Plant Journal* **45**: 83-100
- Joseleau J. P., Imai T., Kuroda K. & Ruel K. (2004) Detection *in situ* and characterization of lignin in the G-layer of tension wood fibres of *Populus deltoides*. *Planta* **219**: 338-345
- Joshi C. P. (2003) Xylem-specific and tension stress-responsive expression of cellulose synthase genes from aspen trees. *Applied Biochemistry and Biotechnology* **105**: 17-26
- Joshi C. P., Bhandari S., Ranjan P., Kalluri U. C., Liang X., Fujino T. & Samuga A. (2004) Genomics of cellulose biosynthesis in poplars. *New phytologist* **164**: 53-61
- Jouanin L., Brasileiro A. C. M., Leple J. C., Pilate G. & Cornu D. (1993) Genetic transformation: a short review of methods and their applications, results and perspectives for forest trees. *Annales des Sciences Forestières* **50**: 325-336
- Jourez B. & Avella-Shaw T. (2003) Effect of gravitational stimulus duration on tension wood formation in young stems of poplar (*P. euramericana* cv 'Ghoy'). *Annals of Forest Science* **60**: 31-41
- Jourez B., Riboux A. & Leclercq A. (2001) Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* cv 'Ghoy'). *IAWA Journal* **22**: 133-157
- Jourez B., Vaianopoulos C. & Hebert J. (2003) Effect of intensity of gravitational stimulus artificially induced on growth and formation of tension wood in young shoots of poplar (*P. euramericana* cv. Ghoy). *Annals of Forest Science* **60**: 789-802
- Justus C., Anderhag P., Goins J. & Lazzaro M. (2004) Microtubules and microfilaments coordinate to direct a fountain streaming pattern in elongating conifer pollen tube tips. *Planta* **219**: 103-109

## Bibliographie

- Kadmas J. L. & Beckerle M. C. (2004) The LIM domain: from the cytoskeleton to the nucleus. *Nature Reviews Molecular Cell Biology* **5**: 920-931
- Kalluri U. C. & Joshi C. P. (2003) Isolation and characterization of a new, full-length cellulose synthase cDNA, *PtrCesA5* from developing xylem of aspen trees. *Journal of Experimental Botany* **54**: 2187-2188
- Kalluri U. C. & Joshi C. P. (2004) Differential expression patterns of two cellulose synthase genes are associated with primary and secondary cell wall development in aspen trees. *Planta* **220**: 47-55
- Kandasamy M. K., McKinney E. C. & Meagher R. B. (2002) Plant profilin isoforms are distinctly regulated in vegetative and reproductive tissues. *Cell Motility and the Cytoskeleton* **52**: 22-32
- Kang K. & Soh W. (1994) Development of reaction tissues in the hypocotyl of *Ricinus communis* L. seedling grown horizontally. *Phytomorphology* **44**: 207-221
- Kao Y.-Y., Harding S. A. & Tsai C.-J. (2002) Differential expression of two distinct phenylalanine ammonia-lyase genes in condensed tannin-accumulating and lignifying cells of quaking aspen. *Plant Physiology* **130**: 796-807
- Kaothien P., Kawaoka A., Ebinuma H., Yoshida K. & Shinmyo A. (2002) Ntlm1, a PAL-box binding factor, controls promoter activity of the horseradish wound-inducible peroxidase gene. *Plant Molecular Biology* **49**: 591-599
- Karlsson O., Thor S., Norberg T., Ohlsson H. & Edlund T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344**: 879-882
- Karpinska B., Karlsson M., Srivastava M., Stenberg A., Schrader J., Sterky F., Bhalerao R. & Wingsle G. (2004) MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen. *Plant Molecular Biology* **56**: 255-270
- Kaul R. B. (1995) Reproductive structure and organogenesis in a cottonwood, *Populus deltoides* (Salicaceae). *International Journal of Plant Sciences* **156**: 172-180
- Kawaoka A. & Ebinuma H. (2001) Transcriptional control of lignin biosynthesis by tobacco LIM protein. *Phytochemistry* **57**: 1149-1157
- Kawaoka A., Kaothien P., Yoshida K., Endo S., Yamada K. & Ebinuma H. (2000) Functional analysis of tobacco LIM protein Ntlm1 involved in lignin biosynthesis. *The Plant Journal* **22**: 289-301
- Kawaoka A., Nanto K., Ishii K. & Ebinuma H. (2006) Reduction of lignin content by suppression of expression of the LIM domain transcription factor in *Eucalyptus camaldulensis*. *Silvae Genetica* **55**: 269-277
- Khurana B., Khurana T., Khaire N. & Noegel A. A. (2002) Functions of LIM proteins in cell polarity and chemotactic motility. *EMBO Journal* **21**: 5331-5342
- Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., *et al.* (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* **301**: 376-379
- Kim S.-J., Kim M.-R., Bedgar D. L., Moinuddin S. G. A., Cardenas C. L., Davin L. B., Kang C. & Lewis N. G. (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 1455-1460
- Kim-Kaneyama J.-r., Suzuki W., Ichikawa K., Ohki T., Kohno Y., Sata M., Nose K. & Shibayama M. (2005) Uni-axial stretching regulates intracellular localization of Hic-5 expressed in smooth-muscle cells *in vivo*. *Journal of Cell Science* **118**: 937-949

## Bibliographie

- Kinzler K. W. & Vogelstein B. (1989) Whole genome PCR: application to the identification of sequences bound by gene regulatory protein. *Nucleic Acids Research* **17**: 3645-3653
- Kirchner J., Forbush K. A. & Bevan M. J. (2001) Identification and characterization of thymus LIM protein: targeted disruption reduces thymus cellularity. *Molecular and Cellular Biology* **21**: 8592-8604
- Knoll R., Hoshijima M., Hoffman H. M., Person V., Lorenzen-Schmidt I., Bang M.-L., Hayashi T., Shiga N., Yasukawa H., Schaper W., *et al.* (2002) The cardiac mechanical stretch sensor machinery involves a Z disc complex that is defective in a subset of Human dilated cardiomyopathy. *Cell* **111**: 943-955
- Ko J.-H., Yang S. H., Park A. H., Lerouxel O. & Han K.-H. (2007) ANAC012, a member of the plant-specific NAC transcription factor family, negatively regulates xylary fiber development in *Arabidopsis thaliana*. *The Plant Journal* **50**: 1035-1048
- Kohler A., Delaruelle C., Martin D., Encelot N. & Martin F. (2003) The poplar root transcriptome: analysis of 7000 expressed sequence tags. *FEBS Letters* **542**: 37-41
- Kohorn B. D. (2000) Plasma membrane-cell wall contacts. *Plant Physiology* **124**: 31-38
- Kong Y., Flick M., Kudla A. & Konieczny S. (1997) Muscle LIM protein promotes myogenesis by enhancing the activity of MyoD. *Molecular and Cellular Biology* **17**: 4750-4760
- Kosarev P., Mayer K. & Hardtke C. (2002) Evaluation and classification of RING-finger domains encoded by the *Arabidopsis* genome. *Genome Biology* **3**: 1-12
- Kovar D. R., Staiger C. J., Weaver E. A. & McCurdy D. W. (2000) AtFim1 is an actin filament crosslinking protein from *Arabidopsis thaliana*. *The Plant Journal* **24**: 625-636
- Kubo M., Udagawa M., Nishikubo N., Horiguchi G., Yamaguchi M., Ito J., Mimura T., Fukuda H. & Demura T. (2005) Transcription switches for protoxylem and metaxylem vessel formation. *Genes and Development* **19**: 1855-1860
- Kurek I., Kawagoe Y., Jacob-Wilk D., Doblin M. & Delmer D. (2002) Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 11109-11114
- Lachaud S., Catesson A. M. & Bonnemain J. L. (1999) Structure and functions of the vascular cambium. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* **322**: 633-650
- Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685
- Lafarguette F., Leplé J.-C., Déjardin A., Laurans F., Costa G., Lesage-Descauses M.-C. & Pilate G. (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New phytologist* **164**: 107-121
- Laitinen R. A. E., Immanen J., Auvinen P., Rudd S., Alatalo E., Paulin L., Ainasoja M., Kotilainen M., Koskela S., Teeri T. H., *et al.* (2005) Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae). *Genome Research* **15**: 475-486
- Lapierre C., Pollet B., Petit-Conil M., Toval G., Romero J., Pilate G., Leple J.-C., Boerjan W., Ferret V., De Nadai V., *et al.* (1999) Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiology* **119**: 153-164

## Bibliographie

- Larson P. R. (1994) The vascular cambium: development and structure. *In* The vascular cambium: development and structure. Springer-Verlag., Berlin, 725 pp
- Le J., El-Assal S. E.-D., Basu D., Saad M. E. & Szymanski D. B. (2003) Requirements for *Arabidopsis* *ATARP2* and *ATARP3* during epidermal development. *Current Biology* **13**: 1341-1347
- Lee D., Meyer K., Chapple C. & Douglas C. J. (1997) Antisense suppression of 4-coumarate:coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. *Plant Cell* **9**: 1985-1998
- Leplé J. C., Brasileiro A. C. M., Michel M. F. & Delmotte F. (1992) Transgenic poplars; expression of chimeric genes using four different constructs. *Plant Cell Reports* **11**: 137-141
- Leskinen E. & Alström-Rapaport C. (1999) Molecular phylogeny of *Salicaceae* and closely related *Flacourtiaceae*: Evidence from 5.8 S, ITS 1 and ITS 2 of the rDNA. *Plant Systematics and Evolution* **215**: 209-227
- Li L., Cheng X. F., Leshkevich J., Umezawa T., Harding S. A. & Chiang V. L. (2001) The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell* **13**: 1567-1586
- Li S., Blanchoin L., Yang Z. & Lord E. M. (2003) The putative *Arabidopsis* Arp2/3 complex controls leaf cell morphogenesis. *Plant Physiology* **132**: 2034-2044
- Li X.-B., Fan X.-P., Wang X.-L., Cai L. & Yang W.-C. (2005) The cotton *ACTIN1* gene is functionally expressed in fibers and participates in fiber elongation. *Plant Cell* **17**: 859-875
- Liu Z. H. & Ger M. J. (1997) Changes of enzyme activity during pollen germination in maize, and possible evidence of lignin synthesis. *Australian Journal of Plant Physiology* **24**: 329-335
- Loopstra C. A. & Sederoff R. R. (1995) Xylem-specific gene expression in loblolly pine. *Plant Molecular Biology* **27**: 277-291
- Louis H. A., Pino J. D., Schmeichel K. L., Pomies P. & Beckerle M. C. (1997) Comparison of three members of the cysteine-rich protein family reveals functional conservation and divergent patterns of gene expression. *Journal of Biological Chemistry* **272**: 27484-27491
- Luo M., Xiao Y., Hou L., Luo X., Li D. & Pei Y. (2003) Cloning and expression analysis of a LIM-domain protein gene from cotton (*Gossypium hirsutum* L.). *Acta Genetica Sinica* **30**: 175-182
- Lynch M. & Conery J. S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151-1155
- Ma J., Przibilla E., Hu J., Bogorad L. & Ptashne M. (1988) Yeast activators stimulate plant gene expression. *Nature* **334**: 631-633
- Martin C. & Paz-Ares J. (1997) MYB transcription factors in plants. *Trends in Genetics* **13**: 67-73
- Mathur J., Mathur N., Kernebeck B. & Hulskamp M. (2003a) Mutations in Actin-Related Proteins 2 and 3 affect cell shape development in *Arabidopsis*. *Plant Cell* **15**: 1632-1645
- Mathur J., Mathur N., Kirik V., Kernebeck B., Srinivas B. P. & Hulskamp M. (2003b) *Arabidopsis* *CROOKED* encodes for the smallest subunit of the ARP2/3 complex and controls cell shape by region specific fine F-actin formation. *Development* **130**: 3137-3146
- Matsuzaki J., Masumori M. & Tange T. (2007) Phototropic bending of non-elongating and radially growing woody stems results from asymmetrical xylem formation. *Plant, Cell & Environment* **30**: 646-653



## Bibliographie

- McCarty D. R., Hattori T., Carson C. B., Vasil V., Lazar M. & Vasil I. K. (1991) The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* **66**: 895-905
- Mellerowicz E. J., Baucher M., Sundberg B. & Boerjan W. (2001) Unravelling cell wall formation in the woody dicot stem. *Plant Molecular Biology* **47**: 239-274
- Meyer K., Shirley A. M., Cusumano J. C., Bell-Lelong D. A. & Chapple C. (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 6619-6623
- Michelsen J., Schmeichel K., Beckerle M. & Winge D. (1993) The LIM motif defines a specific zinc-binding protein domain. *Proceedings of the National Academy of Sciences of the United States of America* **90**: 4404-4408
- Michelsen J., Sewell A., Louis H., Olsen J., Davis D., Winge D. & Beckerle M. (1994) Mutational analysis of the metal sites in an LIM domain. *Journal of Biological Chemistry* **269**: 11108-11113
- Miller R. T., Christoffels A. G., Gopalakrishnan C., Burke J., Ptitsyn A. A., Broveak T. R. & Hide W. A. (1999) A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base. *Genome Research* **9**: 1143-1155
- Ministère de l'Agriculture et de la Pêche (2006) La forêt et les industries du bois 2006. *collection Agreste GraphAgri*: 180 pages
- Miroux B. & Walker J. E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* **260**: 289-298
- Mitsuda N., Iwase A., Yamamoto H., Yoshida M., Seki M., Shinozaki K. & Ohme-Takagi M. (2007) NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. *Plant Cell* **19**: 270-280
- Mitsuda N., Seki M., Shinozaki K. & Ohme-Takagi M. (2005) The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell* **17**: 2993-3006
- Mo Y., Nagel C. & Taylor L. (1992) Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 7213-7217
- Mohapatra B., Jimenez S., Lin J. H., Bowles K. R., Coveler K. J., Marx J. G., Chrisco M. A., Murphy R. T., Lurie P. R., Schwartz R. J., *et al.* (2003) Mutations in the muscle LIM protein and alpha-actinin-2 genes in dilated cardiomyopathy and endocardial fibroelastosis. *Molecular Genetics and Metabolism* **80**: 207-215
- Moreau C., Aksenov N., Lorenzo M., Segerman B., Funk C., Nilsson P., Jansson S. & Tuominen H. (2005) A genomic approach to investigate developmental cell death in woody tissues of *Populus* trees. *Genome Biology* **6**: R34
- Mundel C., Baltz R., Eliasson A., Bronner R., Grass N., Kräuter R., Evrard J.-L. & Steinmetz A. (2000) A LIM-domain protein from sunflower is localized to the cytoplasm and/or nucleus in a wide variety of tissues and is associated with the phragmoplast in dividing cells. *Plant Molecular Biology* **42**: 291-302
- Murashige T. & Skoog F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* **15**: 473-497
- Murphy A. & Taiz L. (1999) Naphthylphthalamic acid is enzymatically hydrolyzed at the hypocotyl-root transition zone and other tissues of *Arabidopsis thaliana* seedlings. *Plant Physiology and Biochemistry* **37**: 413-430

## Bibliographie

- Nairn C. J. & Haselkorn T. (2005) Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms. *New phytologist* **166**: 907-915
- Neale D. B. & Savolainen O. (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* **9**: 325-330
- Nieminen K. M., Kauppinen L. & Helariutta Y. (2004) A weed for wood? *Arabidopsis* as a genetic model for xylem development. *Plant Physiology* **135**: 653-659
- Nishiya N., Sabe H., Nose K. & Shibanuma M. (1998) The LIM domains of hic-5 protein recognize specific DNA fragments in a zinc-dependent manner *in vitro*. *Nucleic Acids Research* **26**: 4267-4273
- Nix D. A. & Beckerle M. C. (1997) Nuclear-cytoplasmic shuttling of the focal contact protein, zyxin: A potential mechanism for communication between sites of cell adhesion and the nucleus. *Journal of Cell Biology* **138**: 1139-1147
- Ohl S., Hedrick S. A., Chory J. & Lamb C. J. (1990) Functional properties of a phenylalanine ammonia-lyase promoter from *Arabidopsis*. *Plant Cell* **2**: 837-848
- Okuyama T., Yamamoto H., Yoshida M., Hattori Y. & Archer R. R. (1994) Growth stresses in tension wood: role of microfibrils and lignification. *Annales des Sciences Forestieres* **51**: 291-300
- Ooka H., Satoh K., Doi K., Nagata T., Otomo Y., Murakami K., Matsubara K., Osato N., Kawai J., Carninci P., *et al.* (2003) Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research* **10**: 239-247
- Page R. D. M. (1996) Tree View: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358
- Palmieri M. & Kiss J. Z. (2005) Disruption of the F-actin cytoskeleton limits statolith movement in *Arabidopsis* hypocotyls. *Journal of Experimental Botany* **56**: 2539-2550
- Panstruga R. (2005) Discovery of novel conserved peptide domains by ortholog comparison within plant multi-protein families. *Plant Molecular Biology* **59**: 485-500
- Paredez A. R., Somerville C. R. & Ehrhardt D. W. (2006) Visualization of cellulose synthase demonstrates functional association with microtubules. *Science* **312**: 1491-1495
- Patzlaff A., McInnis S., Courtenay A., Surman C., Newman L. J., Smith C., Bevan M. W., Mansfield S., Whetten R. W., Sederoff R. R., *et al.* (2003) Characterisation of a pine MYB that regulates lignification. *The Plant Journal* **36**: 743-754
- Paux E., Carocha V., Marques C., Mendes de Sousa A., Borralho N., Sivadon P. & Grima-Pettenati J. (2005) Transcript profiling of *Eucalyptus* xylem genes during tension wood formation. *New phytologist* **167**: 89-100
- Paux E., Tamasloukht M. B., Ladouce N., Sivadon P. & Grima-Pettenati J. (2004) Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant Molecular Biology* **55**: 263-280
- Pear J. R., Kawagoe Y., Schreckengost W. E., Delmer D. P. & Stalker D. M. (1996) Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 12637-12642
- Peer W. A., Brown D. E., Tague B. W., Muday G. K., Taiz L. & Murphy A. S. (2001) Flavonoid accumulation patterns of transparent testa mutants of *Arabidopsis*. *Plant Physiology* **126**: 536-548
- Peng L., Kawagoe Y., Hogan P. & Delmer D. (2002) Sitosterol-beta -glucoside as primer for cellulose synthesis in plants. *Science* **295**: 147-150
- Perez-Alvarado G. C., Miles C., Michelsen J. W., Louis H. A., Winge D. R., Beckerle M. C. & Summers M. F. (1994) Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP. *Nature Structural Biology* **1**: 388-398

## Bibliographie

- Philippe G., Baldet B., Héois B. & Ginisty C. (2006) Reproduction sexuée des conifères et production de semences en vergers à graines. *Collection Synthèses, CEMAGREF*: 570 pp
- Pilate G., Chabbert B., Cathala B., Yoshinaga A., Leplé J.-C., Laurans F., Lapierre C. & Ruel K. (2004a) Lignification and tension wood. *Comptes Rendus Biologies* **327**: 889-901
- Pilate G., Déjardin A., Laurans F. & Leplé J.-C. (2004b) Tension wood as a model for functional genomics of wood formation. *New phytologist* **164**: 63-72
- Plomion C., Lalanne C., Claverol S., Meddour H., Kohler A., Bogeat-Triboulot M.-B., Barre A., Le Provost G., Dumazet H., Jacob D., *et al.* (2006) Mapping the proteome of poplar and application to the discovery of drought-stress responsive proteins. *PROTEOMICS* **6**: 6509-6527
- Plomion C., Leprovost G. & Stokes A. (2001) Wood formation in trees. *Plant Physiology* **127**: 1513-1523
- Pomies P., Louis H. A. & Beckerle M. C. (1997) CRP1, a LIM domain protein implicated in muscle differentiation, interacts with alpha-actinin. *Journal of Cell Biology* **139**: 157-168
- Prassinis C., Ko J.-H., Yang J. & Han K.-H. (2005) Transcriptome profiling of vertical stem segments provides insights into the genetic regulation of secondary growth in hybrid aspen trees. *Plant and Cell Physiology* **46**: 1213-1225
- Prassler J., Murr A., Stocker S., Faix J., Murphy J. & Marriott G. (1998) DdLIM is a cytoskeleton-associated protein involved in the protrusion of lamellipodia in *Dictyostelium*. *Molecular Biology of the Cell* **9**: 545-559
- Raes J., Rohde A., Christensen J. H., Van de Peer Y. & Boerjan W. (2003) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiology* **133**: 1051-1071
- Ralph J., Lapierre C., Marita J. M., Kim H., Lu F., Hatfield R. D., Ralph S., Chapple C., Franke R., Hemm M. R., *et al.* (2001) Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. *Phytochemistry* **57**: 993-1003
- Ralph S., Oddy C., Cooper D., Yueh H., Jancsik S., Kolosova N., Philippe R. N., Aeschliman D., White R., Huber D., *et al.* (2006) Genomics of hybrid poplar (*Populus trichocarpa* x *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. *Molecular Ecology* **15**: 1275-1297
- Ramachandran S., Christensen H. E. M., Ishimaru Y., Dong C.-H., Chao-Ming W., Cleary A. L. & Chua N.-H. (2000) Profilin plays a role in cell elongation, cell shape maintenance, and flowering in *Arabidopsis*. *Plant Physiology* **124**: 1637-1647
- Ranocha P., Chabannes M., Chamayou S., Danoun S., Jauneau A., Boudet A.-M. & Goffner D. (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiology* **129**: 145-155
- Reinhard M., Zumbunn J., Jaquemar D., Kuhn M., Walter U. & Trueb B. (1999) An alpha - actinin binding site of zyxin is essential for subcellular zyxin localization and alpha - actinin recruitment. *Journal of Biological Chemistry* **274**: 13410-13418
- Ren H. & Xiang Y. (2007) The function of actin-binding proteins in pollen tube growth. *Protoplasma* **230**: 171-182
- Richmond T. (2000) Higher plant cellulose synthases. *Genome Biology* **1**: 1-6
- Riechmann J. L., Heard J., Martin G., Reuber L., -Z. C., Jiang, Keddie J., Adam L., Pineda O., Ratcliffe O. J., *et al.* (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**: 2105-2110

## Bibliographie

- Rival A., Beule T., Barre P., Hamon S., Duval Y. & Noirot M. (1997) Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq) tissue cultures and seed-derived plants. *Plant Cell Reports* **16**: 884-887
- Rogers L. A. & Campbell M. M. (2004) The genetic control of lignin deposition during plant growth and development. *New phytologist* **164**: 17-30
- Rogers L. A., Dubos C., Surman C., Willment J., Cullis I. F., Mansfield S. D. & Campbell M. M. (2005) Comparison of lignin deposition in three ectopic lignification mutants. *New phytologist* **168**: 123-140
- Rohde A., Morreel K., Ralph J., Goeminne G., Hostyn V., De Rycke R., Kushnir S., Van Doorselaere J., Joseleau J.-P., Vuylsteke M., *et al.* (2004) Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell* **16**: 2749-2771
- Rohman M. & Harrison-Lavoie K. J. (2000) Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. *Protein Expression and Purification* **20**: 45-47
- Romagnoli S., Cai G. & Cresti M. (2003) *In vitro* assays demonstrate that pollen tube organelles use kinesin-related motor proteins to move along microtubules. *Plant Cell* **15**: 251-269
- Roudier F., Fernandez A. G., Fujita M., Himmelspach R., Borner G. H. H., Schindelman G., Song S., Baskin T. I., Dupree P., Wasteneys G. O., *et al.* (2005) COBRA, an *Arabidopsis* extracellular glycosyl-phosphatidyl inositol-anchored protein, specifically controls highly anisotropic expansion through its involvement in cellulose microfibril orientation. *Plant Cell* **17**: 1749-1763
- Roy S., Jauh G. Y., Hepler P. K. & Lord E. M. (1998) Effects of Yariv phenylglycoside on cell wall assembly in the lily pollen tube. *Planta* **204**: 450-458
- Ruan Y. (2007) Rapid cell expansion and cellulose synthesis regulated by plasmodesmata and sugar: insights from the single-celled cotton fibre. *Functional Plant Biology* **34**: 1-10
- Sabatini S., Beis D., Wolkenfelt H., Murfett J., Guilfoyle T., Malamy J., Benfey P., Leyser O., Bechtold N., Weisbeek P., *et al.* (1999) An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99**: 463-472
- Sadler I., Crawford A., Michelsen J. & Beckerle M. (1992) Zyxin and cCRP: two interactive LIM domain proteins associated with the cytoskeleton. *Journal of Cell Biology* **119**: 1573-1587
- Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406-425
- Samuga A. & Joshi C. P. (2002) A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to *Arabidopsis AtCesA7 (irx3)* gene associated with secondary cell wall synthesis. *Gene* **296**: 37-44
- Samuga A. & Joshi C. P. (2004) Differential expression patterns of two new primary cell wall-related cellulose synthase cDNAs, *PtrCesA6* and *PtrCesA7* from aspen trees. *Gene* **334**: 73-82
- Sanderson M. J., Thorne J. L., Wikstrom N. & Bremer K. (2004) Molecular evidence on plant divergence times. *American Journal of Botany* **91**: 1656-1665
- Sardar H. S., Yang J. & Showalter A. M. (2006) Molecular interactions of arabinogalactan proteins with cortical microtubules and F-Actin in bright yellow-2 tobacco cultured cells. *Plant Physiology* **142**: 1469-1479
- Schindelman G., Morikami A., Jung J., Baskin T. I., Carpita N. C., Derbyshire P., McCann M. C. & Benfey P. N. (2001) COBRA encodes a putative GPI-anchored protein, which

## Bibliographie

- is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes & Development* **15**: 1115-1127
- Schmeichel K. & Beckerle M. (1997) Molecular dissection of a LIM domain. *Molecular Biology of the Cell* **8**: 219-230
- Schmeichel K. L. & Beckerle M. C. (1994) The LIM domain is a modular protein-binding interface. *Cell* **79**: 211-219
- Schmeichel K. L. & Beckerle M. C. (1998) LIM domains of cysteine-rich protein 1 (CRP1) are essential for its zyxin-binding function. *Biochemical Journal* **331**: 885-892
- Schrader J., Moyle R., Bhalerao R., Hertzberg M., Lundeberg J., Nilsson P. & Bhalerao R. P. (2004a) Cambial meristem dormancy in trees involves extensive remodelling of the transcriptome. *The Plant Journal* **40**: 173-187
- Schrader J., Nilsson J., Mellerowicz E., Berglund A., Nilsson P., Hertzberg M. & Sandberg G. (2004b) A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* **16**: 2278-2292
- Scurfield G. (1973) Reaction wood: its structure and function. *Science* **179**: 647-655
- Sequencing Project International Rice Genome (2005) The map-based sequence of the rice genome. *Nature* **436**: 793-800
- Showalter A. M. (2001) Arabinogalactan-proteins: structure, expression and function. *Cellular and Molecular Life Sciences* **58**: 1399-1417
- Sibout R., Eudes A., Mouille G., Pollet B., Lapierre C., Jouanin L. & Seguin A. (2005) *CINNAMYL ALCOHOL DEHYDROGENASE-C* and *-D* are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell* **17**: 2059-2076
- Sibout R., Eudes A., Pollet B., Goujon T., Mila I., Granier F., Seguin A., Lapierre C. & Jouanin L. (2003) Expression pattern of two paralogs encoding cinnamyl alcohol dehydrogenases in *Arabidopsis*. Isolation and characterization of the corresponding mutants. *Plant Physiology* **132**: 848-860
- Smith D. B. & Johnson K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**: 31-40
- Somerville C. (2006) Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology* **22**: 53-78
- Sonnhammer E. L. L. & Koonin E. V. (2002) Orthology, paralogy and proposed classification for paralog subtypes. *Trends in Genetics* **18**: 619-620
- Southworth D. (1973) Cytochemical reactivity of pollen walls. *J. Histochem. Cytochem.* **21**: 73-80
- Steer M. W. & Steer J. M. (1989) Pollen tube tip growth. *New phytologist* **111**: 323-358
- Steinmetz A., Hoffmann C., Dieterle M., Gatti S., Friederich E., Shen W.-H. & Thomas C. (2006) LIM domain proteins : a novel family of plant actin binding proteins. *ASPB Annual Meeting. Boston, USA*
- Sterck L., Rombauts S., Jansson S., Sterky F., Rouze P. & Van de Peer Y. (2005) EST data suggest that poplar is an ancient polyploid. *New phytologist* **167**: 165-170
- Sterky F., Bhalerao R. R., Unneberg P., Segerman B., Nilsson P., Brunner A. M., Charbonnel-Campaa L., Lindvall J. J., Tandré K., Strauss S. H., *et al.* (2004) A *Populus* EST resource for plant functional genomics. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 13951-13956
- Sterky F., Regan S., Karlsson J., Hertzberg M., Rohde A., Holmberg A., Amini B., Bhalerao R., Larsson M., Villarroel R., *et al.* (1998) Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 13330-13335

## Bibliographie

- Stronach B., Siegrist S. & Beckerle M. (1996) Two muscle-specific LIM proteins in *Drosophila*. *Journal of Cell Biology* **134**: 1179-1195
- Sweetman J., Spurr C., Eliasson A., Gass N., Steinmetz A. & Twell D. (2000) Isolation and characterisation of two pollen-specific LIM domain protein cDNAs from *Nicotiana tabacum*. *Sexual Plant Reproduction* **12**: 339-345
- Szymanski D. B. (2005) Breaking the WAVE complex: the point of *Arabidopsis* trichomes. *Current Opinion in Plant Biology* **8**: 103-112
- Takatsuji H. (1998) Zinc-finger transcription factors in plants. *Cellular and Molecular Life Sciences* **54**: 582-596
- Takeuchi M., Déjardin A., Leplé J.-C., Laurans F., Millet N., Aguié-Béghin V., Chabbert B. & Pilate G. (2007) Unravelling the function of fasciclin-like arabinogalactan proteins in tension wood formation. *XIth Cell Wall Meeting, Copenhagen, Denmark*
- Tamagnone L., Merida A., Parr A., Mackay S., Culianez-Macia F. A., Roberts K. & Martin C. (1998) The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *Plant Cell* **10**: 135-154
- Taylor L. P. & Hepler P. K. (1997) Pollen germination and tube growth. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**: 461-491
- Taylor N. G., Howells R. M., Huttly A. K., Vickers K. & Turner S. R. (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 1450-1455
- Taylor N. G., Laurie S. & Turner S. R. (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. *Plant Cell* **12**: 2529-2540
- Taylor N. G., Scheible W.-R., Cutler S., Somerville C. R. & Turner S. R. (1999) The *irregular xylem3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell* **11**: 769-780
- Telewski F. W. (2006) A unified hypothesis of mechanoperception in plants. *American Journal of Botany* **93**: 1466-1476
- Thain A., Gaston K., Jenkins O. & Clarke A. R. (1996) A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends in Genetics* **12**: 209-210
- The Angiosperm Phylogeny Group (2003) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* **141**: 399-436
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815
- Thomas C., Hoffmann C., Dieterle M., Van Troys M., Ampe C. & Steinmetz A. (2006) Tobacco WLIM1 is a novel F-Actin binding protein involved in actin cytoskeleton remodeling. *Plant Cell* **18**: 2194-2206
- Thomas C., Hoffmann C., Gatti S. & Steinmetz A. (2007) LIM proteins, a novel class of actin cytoskeleton organizers in plants. *Plant Signaling and Behavior* **2**: 99-100
- Thumma B. R., Nolan M. F., Evans R. & Moran G. F. (2005) Polymorphisms in Cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**: 1257-1265
- Timell T. E. (1969) The chemical composition of tension wood. *Svensk papperstidning* **72**: 173-181
- Timell T. E. (1986) Compression wood in Gymnosperms. Bibliography, historical background, determination, structure, chemistry, topochemistry, physical properties, origin, and formation of compression wood. *In* Compression wood in Gymnosperms, Vol 1. Springer-Verlag, Berlin, Germany, 706 pp

## Bibliographie

- Tran T., Singleton C., Fraley T. & Greenwood J. (2005) Cysteine-rich protein 1 (CRP1) regulates actin filament bundling. *BMC Cell Biology* **6**: 45
- Tranbarger T. J., Al-Ghazi Y., Muller B., Teyssendier De La Serve B., Doumas P. & Touraine B. (2003) Transcription factor genes with expression correlated to nitrate-related root plasticity of *Arabidopsis thaliana*. *Plant, Cell and Environment* **26**: 459-469
- Tsai C.-J., Harding S. A., Tschaplinski T. J., Lindroth R. L. & Yuan Y. (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New phytologist* **172**: 47-62
- Turner S., Gallois P. & Brown D. (2007) Tracheary element differentiation. *Annual Review of Plant Biology* **58**: 407-433
- Tuskan G. A., DiFazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A., *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596-1604
- Vidali L. & Hepler P. K. (2001) Actin and pollen tube growth. *Protoplasma* **215**: 64-76
- Vidali L., McKenna S. T. & Hepler P. K. (2001) Actin polymerization is essential for pollen tube growth. *Molecular Biology of the Cell* **12**: 2534-2545
- Villar M., Gaget M. & Dumas C. (1987) The route of the pollen tube from stigma to ovule in *Populus nigra*: a new look. *Annales des Sciences Forestières* **44**: 259-264
- Villar M., Gaget M., Rougier M. & Dumas C. (1993) Pollen-pistil interactions in *Populus*: beta -galactosidase activity associated with pollen-tube growth in the crosses *P. nigra* x *P. nigra* and *P. nigra* x *P. alba*. *Sexual Plant Reproduction* **6**: 249-256
- Wada M., Okano T., Sugiyama J. & Horii F. (1995) Characterization of tension and normally lignified wood cellulose in *Populus maximowiczii*. *Cellulose* **2**: 223-233
- Wang Y. & Gilmore T. D. (2003) Zyxin and paxillin proteins: focal adhesion plaque LIM domain proteins go nuclear. *Biochimica et Biophysica Acta* **1593**: 115-120
- Way J. C. & Chalfie M. (1988) *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54**: 5-16
- Weiskirchen R. & Günther K. (2003) The CRP/MLP/TLP family of LIM domain proteins: acting by connecting. *BioEssays* **25**: 152-162
- Wenham M. W. & Cusick F. (1975) The growth of secondary wood fibres. *New phytologist* **74**: 247-261
- Whetten R., Sun Y.-H., Zhang Y. & Sederoff R. (2001) Functional genomics and cell wall biosynthesis in loblolly pine. *Plant Molecular Biology* **47**: 275-291
- Whetten R. W., MacKay J. J. & Sederoff R. R. (1998) Recent advances in understanding lignin biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**: 585-609
- Wikström N., Savolainen V. & Chase M. W. (2001) Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society B, Biological Sciences* **268**: 2211-2220
- Wu L., Joshi C. P. & Chiang V. L. (2000) A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. *The Plant Journal* **22**: 495-502
- Xu Y. & Uberbacher E. C. (1997) Automated gene identification in large-scale genomic sequences. *Journal of Computational Biology* **4**: 325-338
- Ye Z.-H. (2002) Vascular tissue differentiation and pattern formation in plants. *Annual Review of Plant Biology* **53**: 183-202
- Yokoyama R. & Nishitani K. (2006) Identification and characterization of *Arabidopsis thaliana* genes involved in xylem secondary cell walls. *Journal of Plant Research* **119**: 189-194

## Bibliographie

- Yoshida K., Endo M., Nakazono M., Fukuda H., Demura T., Tsuchiya T. & Watanabe M. (2005) cDNA microarray analysis of gene expression changes during pollination, pollen-tube elongation, fertilization, and early embryogenesis in rice pistils. *Sexual Plant Reproduction* **17**: 269-275
- Yoshida M., Ohta H. & Okuyama T. (2002) Tensile growth stress and lignin distribution in the cell walls of black locust (*Robinia pseudoacacia*). *Journal of Wood Science* **48**: 99-105
- Yoshigi M., Hoffman L. M., Jensen C. C., Yost H. J. & Beckerle M. C. (2005) Mechanical force mobilizes zyxin from focal adhesions to actin filaments and regulates cytoskeletal reinforcement. *Journal of Cell Biology* **171**: 209-215
- Yoshizawa N., Inami A., Miyake S., Ishiguri F. & Yokota S. (2000) Anatomy and lignin distribution of reaction wood in two *Magnolia* species. *Wood Science and Technology* **34**: 183-196
- Yuceer C., Land S. B., Jr., Kubiske M. E. & Harkess R. L. (2003) Shoot morphogenesis associated with flowering in *Populus deltoides* (*Salicaceae*). *American Journal of Botany* **90**: 196-206
- Zhang Y., Brown G., Whetten R., Loopstra C. A., Neale D., Kieliszewski M. J. & Sederoff R. R. (2003) An arabinogalactan protein associated with secondary cell wall formation in differentiating xylem of loblolly pine. *Plant Molecular Biology* **52**: 91-102
- Zhang Y., Sederoff R. R. & Allona I. (2000) Differential expression of genes encoding cell wall proteins in vascular tissues from vertical and bent loblolly pine trees. *Tree Physiology* **20**: 457-466
- Zhong R., Burk D. H., Morrison W. H., III & Ye Z.-H. (2002) A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. *Plant Cell* **14**: 3101-3117
- Zhong R., Demura T. & Ye Z.-H. (2006) SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* **18**: 3158-3170
- Zhong R., Morrison W. H., III, Freshour G. D., Hahn M. G. & Ye Z.-H. (2003) Expression of a mutant form of cellulose synthase *AtCesA7* causes dominant negative effect on cellulose biosynthesis. *Plant Physiology* **132**: 786-795
- Zhong R., Richardson E. & Ye Z.-H. (2007) Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta* **225**: 1603-1611



## Annexes

### Annexe 1. Séquences génomiques corrigées des gènes de *Populus trichocarpa* codant pour des protéines à domaine LIM

Cette annexe montre, pour chaque modèle de gène *PtLIM*, la séquence génomique corrigée, puis les séquences déduites des transcrits, des séquences codantes (CDS), et des protéines. De la même manière que sur le site web du génome de *Populus trichocarpa* ([http://genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html)), les régions non traduites en 5' et 3' sont indiquées par des lettres bleues, et les CDS par des lettres rouges. Les motifs microsatellites sont indiqués par des lettres vertes.

#### PtW LIM1a gène IG\_XIV:598591-600291

```
>PtW LIM1a_grail3.0126006201
GGTTGGTGCATGTATATACTATACTACTTCGTGCCTCGTTGAATTTGCATTTTACATTTTTATAAGGCAGCAAGTCTACTGTTGAACAGTTTA
TGACAGCCCACATACACCCTTGCTTGTGCATTTAACCACATTAATAACACCAGATCCCTTCTTTTTCTTTTTTTTATTAACATGGATTCC
GGGTCACTTTATACGCATCTCAATTAATCTCACGAGCCCTGAAATTAACGACCATGCAAGTCTCCAGGAATTCGAGGTTTGTGGGACTCGAAT
TGATGATCTCTTGAGGACAACTTAGAATCTGACCAGTTGAGTTACACCTCTTAGAGTTAGATCCCTTGATTTTTTAGTTTATAGTAAATTTGAT
GCAGAAAATTTATTTTGTATATTTCTAGAAATTCGGTAGCTTATTAACACTATTTAAAAATAATTTATAAATTTTAAATGGGTGCTAAACTA
ATCTATCTAATTAAGATCAACTTAAATTTAAAAATTTGAAAACCTAACTAGAAAAGAAATATGAAAATAACTAGAAAAGAAATAGAAAATCCCAAG
CAACAATTTCTAAGCTTAAATTTAAACATGAAAATGATCTCGACCCATCATAGAATTGAGTTTATGAAACTTTTTAAGAAAATATGAGTCTGATC
CAATAATAAATTAATAATTACACATGTTTTTGTGATAAATTTCTTACATTAATTAATGCATATCAATACGGCTCATGAATTTTGGCTAAATA
TTGCCATTTGGCTATTATGCCTTTGACATCACAAAATAAAGGAAAATCTCGACAAGACATTTACCCAAATGAGGGTAAACATTTGGAGAAAA
TCCAACCACACCACAAAACAAAGTTAGGATTTAGATTTTCCACGGTTTATTACAGAGTTAAAGAGTACAGGCAAAAATTAATAAATAAATA
AAATTAATAAATAAATCTATAGACAGCGGTAGTTAGGTGCTTCAAAGTTGGACAAAATGATTGCCAATTGCCACGAACCTCAATTATACAACACTG
CTTTTGGTCTCAGCAGCATCCTGCAGCCCACTACACCTGTTTCATCATTTCTCTTCCCTCATCTCCCTCATCACTACTCTCCACATCAGTCCC
TAAACAGAAAACAAAATAAAGGGAGGAGATAGAAGCATAGAGACTAAAGTTGCATCTGGGTGAGAATTCGAAGCCATGGCATTTCGAGGAACAA
CCAGAAGTGATGGCATGTGACAAGACTGTCTATCTTTGTCGACAAGTTAACAGCCGATAACCGTGCCTACCATTAAGGCTTGCTTCCGATGCCA
TCATTGCAAAGGAACCCCTCAAGGTATGCTGTAAGAATTTTGATAAGAGATCAAAAACCTGGTAGATCCATGTATTTTAGCAAATGTTAAAGTTG
GATGCCTAGTTGTTTTACTAAATACAGATTGAGCTCCATGTATTTCCAGCATTTCTGTAGCTTCTTGGTTCTAGTTTTGTCCCTCCCGGTCTTA
GCTATCCACCCTTCATAATTGCATACCCGACCTCGACTTGCTCTGTATCATGGATGACTGGATTATCATGTTAATGTTCAAATTCAAAACCCCTGC
AATATATGAGCAGTGGCATATCAGTTTTTAGATACAAAACCCGTCAGAAAATTAACAGTGCCTAAAGAAATAGATCAGCCAAACAAATTCGAAATTC
ATTCACGAAAGCAACATTTCAAGGTGGAGGAGCTCCTTATAGGGTACAGTGAAGTATTTTCTACTGTTCTGAAGTGCATGATTTCTTTCAG
CTGTTCTCTTTTGTATGATTTCTCACTGATCTGTTCCACTGTTAAGCGAAGTACAGAGTCTTATTTAGAGAGAGATGGTTGAAGATAATCAATC
AAGTTGATGTTCTTATATGCAATCTGTTGACAGATGTTTCTATTTGCACTGGCAACTACAATTCCTTTGAAGGGGTTCTTACTGCAGGCCA
CATTTTGTATCAACTGTTCAAAGAAGACTGGCAGTCTTGACAAAAGCTTTGAGGGTAAAAGATAAAGAAAATCTTTTAGGAAAGCCAGTCTTCCG
AGTTCTTCATTTCCCTTTTCAAACTGAAATTTACTCGCATTTAACTTCTTTGTCAGGGACACCAAAAATTTGTAAGCCAGAGAAAACCTGTTG
ATGGTAGGTAATTTCTTTTGCCTATATTCCTGACATTTAGATGCTTTCTATGGCTTACAATGAGTTTCTACCACCTCTCTTTGGTTTTGAAAC
ATGATTTGCTTTGAACAGAAACCTGTGTCAACTAAAGTCTCGACCATGTTTCGCTGGAAACAGAGAAAAGTGTTCGGCTGCAAGAACACTGTCT
ATCCAACCTGAGAAGGTAAGATGCCCACATTTCAAAGATACCCCACTCACAACAGCAGCAGCAACTTATTTAGGAAAGCCAGTCTATACTTTCT
TTACTTAAATCATCATATACTAAAATTTTCTGTTAAAAGGTTTCGGTGAATGGAACCTCTTACCACAAAAGCTGCTTCAAATGCATCCACGGAGG
ATGTACAATTAGCCCATCAACTACATTGCACATGAAGGTCGCCTCTACTGCAACACCACCACAACCAACTTATCAAGGAAAAGGGAACTTG
AGCCAACCTGAGGGTGAATTTGAGAAGGACTCCATGAACAACAAAATTAATGGAAGAGAGATTGCTGCCGAGTCATAATCTGCTGAACGATACA
GTTCATCCTGCATTTTCTGCTTATGATGTTGCTGTCAACCCTCATGGTTCAACTTTCCCTGCAGTCCAAATGTCTTTATGCTTATGAGCTA
GATTAATAATCATGATATTTTCATATTTCTCATTGAAGTGTGTGAGCTTTTTTTTTTGTGTTGATTTGGATTTTGTATGCTATTATGGATGCTCG
CATGAGATGAGAAGACTTGATGTTTATAGTATGTGATATGGTGAAGTAAACTCGTGAATTTGTGGATTTCTTATGCTATCTATGACTTCCA
ATAGATGCCGATGATCATCATTATTTGACCCCTTAGAGTCTCTTTGTTATCGCTGGTTTTTTGTCATTTAGATGCAAGGTTCCCAAGTCAAGG
TGGACTTGACTGAACACCCGCAATTTATGCATGCTCATCACTGAATTTAGGGGGCAAGCAAGCAAAATTCAGTATAACTGCTATGCAAGACA
ACGCTATCAAGCTTATAAATTTGGCAAAAGCTGGATTTGACACACAACCTACAATTAATGCTGATCATATGCCACAAGATAAATTTCCAACCT
CCTGTCTCACAATAATCCCTTCAGTGAAGCAGGAGAGAATGTGAAGAGTGTGTGAGAAAAGCATTTCGTATCTCCATCATCTTCCAAGCAAT
ATGGCAGGAAACAAATTTGAAAGATGGCGGAATGCAGGGACGGAAGACAATTTGGTCATACCAATCACCGCAACCATAAAGAAAATCAAATAATAG
GCTAACAGCAACACAAAATTTGGTCAAGTTTCCCATTTCAACTTGAATTTGATTTCCAAAATTTGAAGTTTATCTAATGGCAAATCTTTTACTT
GACAAAGCCACAATAATTTGGATTTGGTGTCTTTCGTTGAAAACATGGCATTTGTTTCTTTCGAAACGGCAATGCAATGCAACAACACTGCGCGGTT
GTAGTGTCTCCCTTACATTCGTAGAAAGCATAGAATTTTTCGATAGTTATCAGACATGGTCTAGCTTGGATCGTGTGTTTGTAAAGTCAATTC
ATAGGTTAGCAGTTCACTCGGCTCTAACCCAGAAAATATAATTTTAAATTTAAAAAAAACCTTAAAAACATTTATTTTAGACAAAATAAAAAA
ATATGAAAGTTAAACCTACCGAATCACCCACTTTACCAGGTCAACTCTTCTCGGAACATCTCTTAATTAGAGACTAGTTCCAGTTGTAAGAGATG
GTGCTTCATATTTAATATGCTTTTTTTTTTGGTACATTAACGTTTGAATATGCTATAAATTCACACTCAAAAACAGAACACAAAGGGA
ATC
```

#### >transcrit\_PtW LIM1a

```
GCCAATTGCCACGAACCTCAATTATACAACACTGCTTTTGGTCTCAGCAGCATCCTGCAGCCCACTACACCTGTTTCATCATTTCTCTTCCCTCAT
CTTCCCTCATCACTACTCTTCCACATCAGTCCCTAAACAGAAAACAAAATAAAGGGAGGAGATAGAAGCATAGAGACTAAAGTTGCATCTGGGT
GAGAATTCGAAGCCATGGCATTTCGAGGAACAACCCAGAAGTGATGGCATGTGACAAGACTGTCTATCTTGTGACAAGTTAACAGCCGATA
```

# Annexes

ACCGTGCCTACCATAAGGCTTGCTTCCGATGCCATCATGCAAGGAACCCCTCAAGCTTGGCAACTACAATTCCTTTGAAGGGGTTCTCTACTG  
CAGGCCACATTTTGATCAACTGTTCAAAGAAGCTGGCAGCTTGCACAAAGCTTTGAGGGGACACCAAAATTTGAAAGCCAGAGAAACCTGTT  
GATGGTGAGAAAACCTGTCTCAACTAAAGTCTCGACCACTTTCGCTGGACCCAGAGAAAAGTGTTCGGCTGCAAGAACACTGCTTATCCAACTG  
AGAAGGTTTCGGTGAATGGAACCTCTTACCACAAAAGCTGCTTCAAATGCATCCAGGAGGATGTACAATTAGCCCATCAAACCTACATTGCACA  
TGAAAGGTCGCCTTACTGCAAAACACCACCACAACCAACTTATCAAGGAAAAGGAAAACCTTGAAGCAACTTGAAGGGTGTATTTGAGAAGGACTCC  
ATGAACAACAAAACCTAAATGGAAGAGAAGTTGCTGCCGAGTCATAAATCTGCTGAACGATACATGTTTCATCTGCATTTTCCTGCCTTATGATGTT  
GCTGTCAACCCTCATGGTTCACACTTTCCTGCAAGTCCAAATGTCTTTATGCTTATGAGCTAGATTAATAATCATGATATTTGATATTTCTCATTTG  
AAGTGTGTGAGCTTTTTTTTTGTTGTTGATTTGGATTTTGTATGCTATTTATGGATGCTCGCATGAGATGAGAAGACTTGTATGGTTATAGTAT  
GTGATATGGTGAAGCTAAACTCGTGAATTTGTGGATTTTATTGCTATCTATTGACTTCCAATAGATGCCGATGATCATATTATTTGACCCCT  
TAGA

>CDS\_PtWLI1a

ATGGCATTTCGAGGAACAACCCAGAAGTGTATGGCATGTGACAAGACTGTCTATCTTGTGCAAGTTAACAGCCGATAACCGTGCCTACCATA  
AGGCTTGGCTCCGATGCCATCATTGCAAGGAACCCCTCAAGCTTGGCAACTACAATTCCTTTGAAGGGGTTCTCTACTGCAGGCCACATTTGTA  
TCAACTGTTCAAAGAAGCTGGCAGCTTGCACAAAGCTTTGAGGGGACCAAAAATTTGAAAGCCAGAGAAAACCTGTTGATGTTGAGAAAACCT  
GTGTCAACTAAAGTCTCGACCATGTTTCGCTGGAACCCAGAGAAAAGTGTTCGGCTGCAAGAACACTGTCTATCCAACTGAGAAGGTTTCGGTGA  
ATGGAACCTCCTTACCACAAAAGCTGCTTCAAATGCATCCAGGAGGATGTACAATTAGCCCATCAAACCTACATTGCACATGAAGGTCGCCTCTA  
CTGCAAAACACCACCACAACCAACTTATCAAGGAAAAGGAAAACCTTGAAGCAACTTGAAGGGTGTATTTGAGAAGGACTCCATGAACAACAAAAC  
AATGGAAGAGAAGTTGCTGCCGAGTCATAA

>PtWLI1a

MAFAGTTQKCMACDKTVYLVDKLTADNRAYHKACFRCHHKGTGLKLGNYNSFEGVLYCRPHFDQLFKRTGSLDKSFEGTPKIVKPEKPVDEKPE  
VSTKVSMTMFAGTREKCFGCKNTVYPTEKVSVNGTTPYHKSCFKCIHGCTISPSNYIAHEGRLYCKHHHNQLIKEKGNLSQLEGDIEKDSMNNKT  
NGREVAAES\*

## PtWLI1b gène LG\_II:8904210-8906209

>PtWLI1b\_estExt\_fgenesh4\_pg.C\_LG\_II1071

ATTTGTGATTTAACACCTCTATACACGGTTATGGTGGAGAGATTTGTACAAACTGTACAGGATCTTTTCTTAACTTACCTTATCAGTAAGC  
TCAAACATCCTCAATGCTTACATTTGAAGAGAACATCGAACCTTTTATTAAGACATCCTTTGCCTTCTAGCTATTTCCCCATCATATTCTTTTT  
CGAGTTGAGGATCTTGTTTCAACAAGAGTGCAAGCCATTCGCACTCTTCTGTTTTTCGCTGTTTGGTCAAATGTCAAAGCACTGGAAAATCA  
TGCAAATTAGGCTCCTGAATTCGTGATCAATTTGAAATTCAGTATACGTGGCATATACAGTTGCTCAACCTTACAGAAATCTAACTAAGCTTC  
CCTGCAGGATCAGTCCAGACTCCAGCTCTCATGGTGCCTCCCTTCTTCTTCAAAGTTAGCCATAGAATAACCTTAATCCCGTGGCCCATTA  
TTTGTATGAAACAGAGATATTTTGTAGGATAATATTAGCAGGGCAGATGTCCAACCTATAACATTTATGGATTTCAACAATAAAGAAAG  
TAATCGAAGCAGGTTGGTACTTGGTATCCTTATACAACGTGCCTGGTTGAATTTGCATATAACATTTTCCATTATGTAGCCAGCCAGGTATATC  
GTCGAACAAATTTATGACAACCCACTTACAGCCATGCCTAATCATACCAGTTCCTATTGTTTTAGTTTGTAGCAACAGTAAAGGAACCTTTTC  
ATAAGACATTTGACCAAAGCAAGGTAATGTCTCTCATGAAAATCCAACCACATTAACAAACTAAAGCTAAATTTTCTATTTTATATGTAT  
AGTATTTATAAAGTTGGATTTAGAGTTACAGGGGAAAATAGAGATAAAAATATATATTGCAACCCCTAGTTAGGCGCTCAAAGTTGGCAAAAATGAT  
TGCCATTTGCCACTAACTTCAATTTGTACAACCTCTGCTTTGGTCTCCGCAATAAACCCTGCTGCACACATTTGTCCATCATTTCTCTCTCTTATCTT  
CCCTCATTACTTTTCTCATTTAGTCCCTCCACAGAAACAAGCAACAAGGGAGGCGATAGGAGCATAGGAGATTACAGTTGCATTTTCTTAAGC  
TGAGAAATTTGAGGCCATGGCAATTTGCAGGAACAACCCAGAAATGTATGGCATGTGACAGACTGTCTATCTTTGTGGCAGGTTAGCTGTGTAT  
AACCGTGTGTACACAAGGCTTGGTTCGATGCCATTTGATGAGGAACCCCTCAAGGTATGCTCTAAATATGCAAGAGTTTAAACAGATTTAA  
AAAACTCGGTAGATCCTATATATTTTAGCACATGTTAAAGTTGAATTCCTAGACGTTTTCTAGTAAATACAGATTTGAGCTCCATTTATCCAG  
TATTGCTTAGCTTCTGGTTCAAATTTTGTCTCTTAGTCTTAGCTATATAACTTTTCAATATAGCATATGCCTGGATTTGTCACGTATCATGGA  
TGATTTGATTTCTCAAGTTAATTTTCGAATTCAGTATTTAGCAGAGACGATATAACCTTTTGGATACTGAAACCCCTCCAGAACTAACAGTGGCA  
TGAATACATCAGTCAATAAGATACAAATTTCAATAACACCCGACTGAAGCAACCTTTTCAGTTGGAGGAGCTCCCTCCATGACATTTGCGT  
GATATTTCCGTAGTGGCCATGATTTTTTTTCTCTCTTTTGTGATGACATTTCTCACTGGTCTATTCAGTGTAAACGAAGACTAGCTCCTTCCAG  
CTTTAAGTCGTAGAACAATCCTGATTTGATCGAATTTATGATCCTAACTAACCAATTTAAATTTGGTTGGGAGCCATGAAATGTAATAACGTCCT  
GCATCAGCTCTTAAATGCAATCTTGGACTTTTTTCAGCAATATGATTTAGCTAGAGATAAATAAGATAAATTTATCAAGTTTCAGCTCTTATTT  
GCAGCTTTGGCAACTACAATTTCTTTGAAGGGGTACTCTACTGCGGCCACATTTTGTATCACTCTTCAAAGAAGCTGGAAGTCTTGACAAAAGC  
TTTGAAGGTAATAAATAAGAAAACCTAATATGAAGCCGATTTGTCACAGCTCATTTTCTTTTTCGAATTTCAAACCTTTCTTCTGCTATTTCA  
ATTTCTTTCCAGGGACACCAAAATTTTAAAACCAGAGAAACCTGTTGATGGGGAGGTAAAAAAAATTTCCCTGTATTCCTGGCATTTAGA  
TGCTTTCTTATGCTTTTCAATTTGAGTCTTGAACCCCTCTCTTTTGTCTATAACCATGATTTGTTTTGACAGAAACCTATATCAACTAAAGTC  
TCGACCATGTTTCGCTGGAACAGAGACAAGTGTTCGGCTGCAAGAACACTGTCTATCCAACCTGAGAAGGTAAGACTGCAACATTTCAAACA  
ATCCAACACACTAAGAGCAGACACTGCTCGTTTTCCATGTTCTATGTTTTTTTAAATCATCAATTTACTAAATTTCAAAGAGATAACTGAGCTGTA  
AGGATCTTGACACCTATCATTGATTTGAGACTATTAATAATTTCTCTGTTAAAAGGTTTCGGTGAATGGAACCTTCTTACCACAAAAGCTGCTTCA  
AATGCATTCATGGAGGGTGTACAATTAGCCCATCCAACCTACATGTCACATGAAGGTCGCTCTACTGCAAAACACCACCACAACCAACTTATCAA  
GGAAAAGGGTAATCTGAGCAACTGAGGGTGTGTTGAGAAGGATTCATGAATGACAAAACCTAATGGAAGAGAAGTTTCTGCTGAATCATAA  
AATGCTGAACAGTGTATTTTCTCGCTTACGATGTTGCTGCAACCCCTATGGTTATTTTACACTTCTGCAAGTCAAATGGCTTTATGCTTAC  
GAGTTAGATTGAAATCATGATATTTATCACCAGAAATGTGTGGGTTTTTGTGTTGTTGATTTGGATTTTGTCTGCTACTATGGATGCTCGCGTG  
AGATGAGAAGACCTATTTGATTCAGCATGTGATCATATGATGAGCTAAACTCGTGTATTTTGGGAATGCTTATATGATCTACTGTGCCCGTC  
TTCAGTTCAACAAGTAAGTAAATAATATTAATAATTTGGTATTTTCACTTTTCACTCGAACTCCACTTATCTTACACCTCTCAAGTCAATTTCA  
CAAAAATCGAACTAAAATCAAGTTCAACAGGGTCTTTTTTCTCCGCTGATTTCAACAAACCCGTTTTTCTGGTTGTTTCACTAGATAGTAGA  
CAGAGACATTTGAAATCTTATTAATTCATTCATGCGTATCATAATTAGATAACGAGGCAATTTGACTATCTTAAGAGAGTCATAATTACTCCTG  
CTGTTTACTGCGCTTGGTTGAAATTTTTTCACTTTGACATTCAGAGCATTGACATAAATCACATTTGCGTGAGCATCTCCAAGAACCATCCGAAT  
GCTTTGTTTTAATTAATAGTTGGATTTCTCTTGTCCGTACGACTCTTAGTCTGATTTTCAACGCTCTGGGAAGACCCCTGGAAGGGTCTGTT  
CCAGTCCATCTCTCGGCAACGACGATGACCCACTCTCTCCACAAGAACACTCGAGCAATCCACCAACACCCGATGGGTTTGGGACTGAAAC  
CCTCAAGACCAACCCCTCAAACCAATCCTTTTCTTAAGTTATAAATTCATTTTACCAGCTTACCTTACCTACATTTATTTTACCAGAAAGT  
TGATGTGATTTAGTACAACCCGGTGTGGGAGGCACTCGATCCTCTGGATTTTCAAGAGCCGTCGGGGATGCACCAGACACCAGCAACGTCG  
AGTGTCTTCCAGCTGATAGATCCTATCTCAACTAAGTCAATTTCCAGGATAAATGAGCTGTAAAAAATAACTCTTCTTGGAGCTTCTGCTGAT  
GTCTCAAACCTCTAATATTTGCCCTCAGTCGCCATATTCAGTTCAAGAATGTTTTTACCCTCTGGAGTCTTCAATTTATGCTGGTTTTTTTCT  
CCTTTTAAATGCAAGATTTCTAAGGCAGGCTGGACTCGACTGAACACCTGCAATTTATGCAAGTCTCATCACTTTAAATTTAGGGGTACAGCCAC  
ACAAGC

>transcrit\_PtWLI1b

# Annexes

TAACCTCAATTTGTACAACCTCTGCTTTGGCTCCCGCAATAACCTGCTGCACACATTTTGTCCATCATTTCTCTTCCCTCATTACTT  
TTCTCATTAGTCTCCACAGAAACAAGCAAAAGGGAGGCGATAGGAGCATAGGAGATTACAGTTGCATTTTCTAAGCTGAGAATTTTGAG  
GCCATGGCATTTCAGGAAACAACCCAGAAATGTATGGCATGTGACAAGACTGTCTATCTTGTGGACAGGTTAGCTGATACCGTGTGTACC  
ACAAGGCTTGCTTCCGATGCCATCATTGCAGAGGAACCCCTCAAGCTTGGCACTACAATTCCTTTGAAGGGGTACTCTACTGCAGGCCACATTT  
TGATCAACTCTTCAAAGAAGCTGGAAGTCTTGACAAAAGCTTTGAGGGGACACCAAAAATTTTAAAACCAGAGAAACCTGTTGATGGGGAGAAA  
CCTATATCAACTAAAGTCTCGACCATGTTTCGCTGGAACCAGAGACAAGTGTTCGGCTGCAAGAACACTGTCTATCCAACCTGAGAAGGTTTCGG  
TGAATGGAACCTTACCACAAAAGCTGCTTCAAATGCATTCATGGAGGGTGTACAATTAGCCCATCCAACCTACATTGCACATGAAGGTCGCT  
CTACTGCAAAACCACCACAACCACTTATCAAGGAAAAGGTAATCTGAGCCAACCTGAGGGTGTGTTGAGAAGGATTCATGAATGACAAA  
ACTAATGGAAGAGAAGTTTCTGCTGAATCATAAATGCCTGAACGAGTGTATTTTCTGCTTACGATGTTGCTGTCACCCCTCATGGTTTATTTA  
CACTTCTGCACTTCAATGGCTTTATGCTTACGAGTTAGATTGAAATCATGATATTTATCACCAGAAATGTGTGGGTTTTTGTGTTGTGATTT  
GGGAAATGCTTATATGATCTACTGTGCCCTCTCAGTTCACAAAAGTAAAGTAAATAATATTTAAAAATATTG

>CDS\_PtWLI1b

ATGGCATTTCAGGAAACAACCCAGAAATGTATGGCATGTGACAAGACTGTCTATCTTGTGGACAGGTTAGCTGCTGATAACCGTGTGTACCACA  
AGGCTTGTCTCCGATGCCATCATTGCAGAGGAACCCCTCAAGCTTGGCACTACAATTCCTTTGAAGGGGTACTCTACTGCAGGCCACATTTGA  
TCAACTCTTCAAAGAAGCTGGAAGTCTTGACAAAAGCTTTGAGGGGACACCAAAAATTTTAAAACCAGAGAAACCTGTTGATGGGGAGAAAACCT  
ATATCAACTAAAGTCTCGACCATGTTTCGCTGGAACCAGAGACAAGTGTTCGGCTGCAAGAACACTGTCTATCCAACCTGAGAAGGTTTCGGTGA  
ATGGAACCTTACCACAAAAGCTGCTCAAATGCATTCATGGAGGGTGTACAATTAGCCCATCCAACCTACATTGCACATGAAGGTCGCTCTA  
CTGCAAAACCACCACAACCACTTATCAAGGAAAAGGTAATCTGAGCCAACCTGAGGGTGTGTTGAGAAGGATTCATGAATGACAAAAC  
AATGGAAGAGAAGTTTCTGCTGAATCATAA

>PtWLI1b

MAFAGTTQKCMACDKTVYLVDRLAADNRVYHKACFRCHHRCRGLKLGNYNSFEGVLYCRPHFDQLFKRTGSLDKSFEGTPKILKPEKPVGDGK  
ISTKVSMTFAGTRDKCFGCKNTVYPTEKVSVNGT SYHKSCFKCIHGCT ISPSNYIAHEGRLYCKHHHNQLIKEKGNLSQLEGDVEKDSMNDKT  
NGREVSaes\*

## PtGLIM1a gène LG\_IX:4795506-4797325

>PtGLIM1a\_grail3.0001068602

TCCAGGTAATAATGCCTCCACTGACATAAAGATGTCGTTTGGCCATATTATAAAAAATATAAATGTAGACGTGAATATGTAATATTCACATGAGC  
TTGCTTGTGCTGTGTTTTGAGCTTTAGCCGGAACACTTGAGATATAAACAGCTGACCTGAGAGTGCCTAGATGATGTTGGTATGATAAACCCAC  
ATCTTCCGACAATACTCCGATTTTATTGAAAAGTCTTGTCAATTAATATGCTGGAGTGTGGCCGAAGACAATTTTACATGAATCCAGTAGCTCA  
CTGGCTCACTAGTTATATGCAGCATCATTAATAAAAAAATTCGATTTGATGCTTAATATAAAAAAGATTATCCATCTAGTGAATCATAATTAAC  
TCATGGAATAAAGTAACCAACGATCCTTAAAAATATGAAAACGAAAGAGAAAAGTAAACAGGCGCCAGCTACCAGCTTCATGATGATTTCTAG  
TTTTAGACCCACCACCTTGTGAAAAGATTAACTCATCTGTTCTGTTCTGTCTACCTAGTAATCATCTTTATATCGAATAATAGCTCCTCC  
TTTTTCTCCGATATAAATGTCCTCATTAACCAACTTCTACCATCTGTTGTTATTTATCATGTATGTTGATTTTCTAGGATTTCCCGTTACTA  
TCTCGAGATGGCAGAAATAAAGGAAATATATTCCTGCTGCGTATGTTTTGAAAGGATAGAAATTCATTTAAATATGAATTTAAATATATTT  
TCTCTCATTTCTATTTCGGTAAACCTAACACAAACACAACTAGATGTAGCAATCAAAATCAATATATGAGAAATCTTTCTAAAACTAAAACTA  
GATGAGATGAGATCTTTAATAAATCTTTGAGGCTAAAAGCTAGAACTGGACCTCCCAAGAGTAAAGAAAGAAAATAGAATGATTTATCTGATAA  
GAAATGCTTAGAATCGGACAAATGATAGCTAAGTGGCCAGACCATTTCTATCAATCATGTACAAAATTTGCTTTTCATGTCCCGCAATTCAC  
TGCAACCTTTGATGCTTTTAACTCCTTCGAACTCATCAGCTACGTTCTTCTGCTGCTATAATAATCGAGCCAAAGACAGAAAACAGAC  
GAGAAATATCACAACTGGTGCACAAGTGTTTAGGTTTCTAGCTCCCATTTTGTCTTCTTCTTCTTCTTCAAGGGAGAACAAGGATATGG  
CAGCCTTTGCAGGAACCTCAACAAAATGCATGGCTTGTGACAAGACTGTGTATCTTGTGATAAGTTAAGTGTGATAACAGGGTTTTTCACAA  
GGCCTGCTTTAGATGCCACCTTGCAGGGGTACCTTCAAGGTAACATAATATCTTTTTTTCATGTTTAAATTTTCTTGATCATTTTTGTTTTAAC  
ATGTTTTATGCACTTCTTTGAACCTGTGATGCTCAGGAAAACACTCTTCTTTTCAGAAATGCTTTTTCATATTTGTTGTTGTTGTTGTTGTTG  
TCCATCAAGTTTTCTTCTGATCAGATCCAACAATCTCTTCTTTGATCAATCTAGTTTCCAAATCTCAAGATTTCAATTTGTTGTTGTTGTTG  
TTTTTTTTTTGATACATGCACTAGCAACATCTTTTCTGGCTAAATCTAAGTTTGGATTTCTGAGAATCATCTTTCATTTTCTTATTTCTTT  
TGAAAATGTTTTCCAATTTCTGTTGCTATGGAAGGATGTTTAAATGCTGGTCTCAAGCCAATTAATTAATCTCATTAATTTCTTGTATGAG  
CACTGTTGACATCTGTTTTATGTTTTGCGATTAAGCAACTACTGTTCTTTTGAAGGGTCTGTTACTGCAGACCTCACTATGACCACTCTT  
CAAGAGAACTGGCAGCTTAGACAAGATTTTGAAGGTAATAAGCAATCAAGCCATTAATAAATTTACACGTTCAAAATCTTCTTAATATCAC  
CAATAGCTTGCCTGTCCTTCTATAACTAATTTCTGTCAGAACCGCAAGATTGTGAAACCAGAAAACCTGTTGATAATGAGGTAATCTC  
AAGATTGGAATGAAAGTATGTTGTTTCTGCTATTTATTTTATTTCAATTTGCTATCCACTAATTAAGTACCCTGAACCTGATGTTCTCATAGA  
ACGCCAGCAAAAGTTTGCATTTGTTGCTGCGCAGGAGAAAATGTTGAGCTGCGCACAGACTGTCTATCCGATCGAGAAGGTAAGAGCCGG  
TGACTTTAGTTATAGAACTCAGAGGACCTCCTTAAAAAATGACCAACTTTTTCTGAAAGTGGTCTTTGATATTTCTTACAGGTTACGGTTAAAT  
GGGACTCCATACCACAGGAGCTGTTTTAAGTGTACCCATGGAGGTTGTACTATAAGCCATCTAATACATTGCACATGAGGGAAAACCTCCACT  
GCAAGCATCACCATATCCAATTTTCAAGGAGAGGGAACCTACAGCTCAGCTTGAAGTGAACGTTGAGAAAAATCCAGTCTGATCATCACTGC  
TGTAGAAATGCTGCTGAATCATAACTCAACTAAACACTACTGTGCACATCTTTAGCTATTGCTTTTCATGTCCTTCTACCATACATGGCAGT  
TCAGTTTCTCATGAAATTTTCT  
GTGGGACTATGATGATCGTTGAATTTTCTGCCAAGATAAAGTAACCTAAGATACTTCTGTCAGACTTTGTTGTTGTTGTTGTTGTTGTTGTTG  
TGTAAGATTTG  
CTGCACCGTGGCATTTCAGAAAAGAGGATACGGTCCCACCTACCCTAAAACGGTTTAAAGCACAAAACATTCATTCAAGATAGATAATGAA  
TTCCACGCATAATGAGGCATGTTTATGCTTATAACTTCTTCAATCCATAGCAGTTTCCACCATGTGGCATGTACAGAGCACAAAATTTCCAGT  
GGGTTGTTTAAAGCCAACAGAAAATGCATCCTCTAATCGCTTAAACAGAAACATGACGATTAATAATACAGCGGCAATACTGAAAGGAAACATA  
CACAAAACCTAAGAACTTTGTTGTTGAAATAAATATAACCCAGTGCATATAACTGTATTTTATGATCTAGCCTAACCATGTATATACAAA  
GGAATGAACTCTAGTCCCTTACCCCTTTCTTCTTCAAAAACCCACAAGCAATAACAGGCTGTACAATGACAAGATAAGCATTTGCTTTCA  
CGTTTCTCATGAACTGCTCAGATTTTCGACTGCACAAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
AATAGAGTATCGTGTGTACCGCACAGACCACATTAATAAGCTTCCATCTCAGGAACCTTTATGGTTTTATTGCTTTTAGTATCTGACCTGC  
AAGTATCAGAAGCTCACCATTTCAAACCTGCAGCAGAAACAGGATCTATCTTCTACTACTGTGTTCCCATAAATATCAGTTAGCTTGTGTTG  
TGGAACTTTCAACATTTGATGCATCTGCTCGGCAGACCACTAATGGTAAAGGAGCTTCAAGGGCAGTCTTCAAGTCTAGCCTCGT  
AAATGATATCATCAAGACTACTGGCAGGCCACTAAGTACTTACAGGATCTGACT

>transcrit\_PtGLIM1a

GTACAAAATTTGCTTTCACTGTCACCCGCAATTCAGTGCACCCCTTTGATGCTTTTAACTCCTTCGAACTCATCAGCTACGTTCTTGTCTCTC  
GCTATAATAATCGAGCCAAAGACAGAAAACAGACGAGAAATATCACAACCTGGTGCACAAGTGTTAGGTTTCTAGCTCCCATTTTGTCTTCT  
TCTTCTTCTTCTTCAAGGGAGAACAAGGATATGGCAGCCTTTGCAGGAACCTCAACAAAATGCATGGCTTGTGACAAGACTGTGTATCTTGTCC

## Annexes

ATAAGTTAACTGCTGATAACAGGGTTTTTCACAAGGCCTGCTTTAGATGCCACCATTGCAGGGGTACCCTCAAGTTAAGCAACTACTGTTCTTT  
TGAAGGGTCTGTACTGCAGACCTCACTATGACCAACTCTTCAAGAGAAGCTGGCAGTCTAGACAAGAGTTTTGAAGGAACGCCGAAGATTGTG  
AAACCAGAAAAAAGCTGTTGATAATGAGAACGCCAGCAAGTTTTCAAGATTGTTTTGCTGGCACCAGAAAAAATGTGTAGCTGCCACAAGACTG  
TCTATCCGATCGAGAAGTTACGGTAAATGGGACTCCATACCACAGGAGCTGTTTTAAGTGTACCCATGGAGGTTGTACTATAAGCCCATCTAA  
CTACATTTGCACATGAGGGAAAACTCCACTGCAAGCATCACCATATCCAACCTTTTCAAGGAGAAGGGAAAACTACAGTCAGCTTGAATGAACGT  
GAGAAAAATCCAGTCGTCAGCATCACTGCTGTAGAAATTTGCTGCTGAATCATAA**CAACTAAACACTACTGTGCCACTCTTTAGCTATTGCTT**  
TTCATGCTCTTCCACAGAACTGGCAGTTTCAGTTTCTCATGGTTTTCTTCTCTCTTTTGTATGTCAAGGATTGGTTTTTCACTTGAAGTGAG  
ATGTGTTGGTTTTCTATGGCTGAATTTTGTGGGACTATGATGTATCGTTGAATTTTCTGCCAAGATAAAGTAACTAAGATACCTCGTGGCA  
CAGTTTTGATCCCTGTTTCCCTCAAAAAATGTAAGATTGTGGTGTCTGAACAATATCATCGTGTCTATTTAGCTGTGAACAATTCAGCAG  
GAAAAAGTTGCGTCAACCTAGGAAGA

>CDS\_PtGLIM1a

ATGGCAGCCTTTGCAGGAAGTCAACAAAAATGCATGGCTTGTGACAAGACTGTGTATCTTGTGCGATAAGTTAACTGCTGATAACAGGGTTTTTC  
ACAAGGCCTGCTTTAGATGCCACCATTGCAGGGGTACCCTCAAGTTAAGCAACTACTGTTCTTTTGAAGGGGTTCTGTACTGCAGACCTCACTA  
TGACCAACTCTTCAAGGAAGCTGGCAGTCTAGACAAGATTTTGAAGGAACGCCGAAGATTGTGAAACCAGAAAAAAGCTTGTGTAATGAGGAA  
GCCAGCAAAGTTTTCAAGTTTGTGCTGGCACCAGAAAAATGTGTAGGCTGGCACAAGACTGTCTATCCGATCGAGAAGGTTACGGTAAATG  
GGACTCCATACCACAGGAGCTGTTTTAAGTGTACCCATGGAGGTTGTACTATAAGCCCATCTAACTACATTGCACATGAGGGAAAACTCCACTG  
CAAGCATCACCATATCCAACCTTTTCAAGGAGAAGGGAAAACTACAGTCAGCTTGAATGAACGTGAGAAAAATCCAGTCGTCAGCATCACTGCT  
GTAGAAATGCTGCTGAATCATAA

>PtGLIM1a

MAAFAGTQQKCMACDKTVYLVDKLTADNRVFKACFRCHHCRGTLKLSNYCSFEGVLYCRPHYDQLFKRGTSLDKSFEQTPKIVKPEKPVNDEN  
ASKVSNLFAQRETKCVGCDKTVPIEKVTVVNGTPYHRSCFKCTHGGCTISPSNYIAHEGKHLHCKHHHIQLFKEKGNYSQLENERKKNPVVSITA  
VEIAAES\*

**PtGLIM1b gène LG\_I:27294474-27295832**

>PtGLIM1b\_fgenesh4\_pm.C\_LG\_I001083

TTACGTTCTTGGTCTGAGGTTTGGCAGCTTACATGATTTAAATGAAAAGGGTATTTCTAGCAAAATGCCCTCGTGGTTTTGCCATGAATAAA  
TTTGAGGTGAAATATGAAATCGACATGATTTACAATCTGATCTGATATGATAAAACCACAATTTCCAATTATAATGTTGAGTGTGGTTGAAGAC  
AAATTTGCATTAACGCAAGAGCTCACTGGGTCCTAGTTACATGGAGCATCATGAAAGGCTTTGATTGATGCCAATTACCTAATTTAAAAAAA  
GATTTATCATCTGATGAATCATTGCACTGGCCGATTATAACTCATGGAAGAAAGTAAATCATAGAAATAAACATGACAACGAAAAAGAAACGCTCT  
AGCAAGGCATCCAGCCAGGAAGCTTTCATGATGATATTTAGCACTCAATATAATTTAGACCCACCACCTTTGTGAAAAAGCAATTTACTCATCAT  
GTTCCGCTGTGATCTAGCAAGAAATCACTTTATATCATTAATAACTTCTCTGTTTTTATTTTCCGTAAGTAAAGCAAGCAATTTCT  
AAGAAACAGGAAAGTAGAAAAATCTGGGAGCCAAAAGCTTGAATGTCAACTTCTTTTTAATGGCATGTGTTTTATAATTGTACCTCCAGAAA  
TAGAGAAAGAAAGTAGAATGATTACAACAGAAAGGATTATCCTATCTGATATGAATTAAGTAAATCGGACAAAATGATAGCTAGGTGGCGCCA  
GGCCATTCCTGTCAATCATGTAAGTATTTGCTTTCATGTCAACCCGCTATCTGCAACCCATTTGCTTTTTATCTCTCTCTTTAAGTCTCCCTA  
CTTTCCGTAATAATAGCAAGATACAGCTTGAAGAGAGACAGGACAGTAAGGCAAGGCTAAGGACACAAACAGCAAGAAATACACAATC  
GGTGCACAAGTGTAAAGGTTCTTGGCTCTTATCTCTTTCATCTTTGAGGGGGGAAAAAGAATATGGCAGCCTTTGCAGGAACCCAGCAAAAATGTA  
TGGCTTGTGACAAGACAGTGTATCTTGTGATAAGTTAACGGCTGATAACAGGATTTATCACAAGGCCTGCTTTAGATGCCACCATTGCAGGGG  
TACCCTTAAGTTAACAGCATATCGGTAATATCCTTGTTCATGTTTTATCCATCTCTTTGAAGTGAATGTCTCATGAAAAACCTTTTTTC  
TTCTCAGAAATTTGCTTTATATATTTTTTTTTTTGCTAAAACCTTTCCATCAATTTTTCTTTCTGATCAGATGACACCATTTCTCTTAT  
TAATCTAAGTTCCAAATTTTCAAGGCCCAATTAATTTTTTTTTATATATATACATATGGCCAGCAGCATAAAAACATTTTTAGCGGTAATTAAGA  
GCCAGGATTTCTGAGAATCATCTTTCAATTTCTTGGTCTTTTTGAAATGCTTTCCGTTTTCTGTTGATGGAAGGAATGTCAAATGCTGTCC  
TCAAAGCCAAATTAACCTACTCATTTGCATTTCTGATGAGCATCATCGGCATCACTGTTTTGCTAGCAGCTAAGCAACTCAGTTCTTTTGAAG  
GGTCTGTACTGCAGACCTCACTATGACCAACTCTTCAAGAGAAGCTGGCAGTCTAGACAAGAGTTTCAAGGTTAAAGCAAGTAAAGCCATT  
ATCAAAAATTTACACACTTCAAACTTACTATTTATCTACTAGCCAGTTAAATTTCTTTAACTTAAATTCAAATTTGCCCTCTCCTTTCTATAA  
TTAATTTCTGCCAGAACACCAAAAATTTGTAAGCCAGAAAAAAGCTGTTGATAATGAGGTGATCTCTTAAGTAAATGATTGGAAGATGATAGTA  
TGTGTTTTCTTGTCTATTTATTTTATTTCAATTTCTATCCACAAATAAATAAAGTAACTGATATGTCTCACAGAACGCAAGCAAGTCTCGAAT  
TTGTTTGTGCTGCCAGCAGAAAAATGTTGGTGCACAAGACTGCTATCTATCGAGAAGTAAGAAATGAAGACTTTATTCATAGAAACA  
GAGGATCATCTCTTCAAGAGTACCAACATTTCTGAAAGATTGGTCTTTGATCTTCTTCCAGGTTACAGTAAATGGGACTCCTATACCCTCGGA  
GCTGTTTTAAGTGTACCCATGGAGGTTGTACTATAGCCCATCCAACCTACATTGCACATGAGGGAAAACTCTACTGCAAGCATCACCATATCCA  
ACTTTTCAAGGAGAAAGGGAAGTACAGCCAGCTGAGAAATGAACGCGAGAAAAATCCAGTCACTGTGAGTATCACCCGCTGTGGAATTTGCCGCT  
GAATCATAAATCAACTAAAGACTTGTCTCCCGTCTCTAGCCCTTTTCAGTCTCTCTACCCTAAATGGCATCGCAGGCTCTTTCATGATTTCT  
TTCTCTCTATAATGTAAGGATTGGTTTTCTGCTTGCAGATTTGATGTTGTTAGTCTTCCATGGCTGTAATTTTGGGAACTCTGAGTGTGATG  
TGTGAGTTGAAATTTCTGCCAAGATTGAGTAACCTGAGATACTAGTGGCAGTGTGATCTCTGTTTCTCAAATACGTAAGATTTCGG  
TATCTCTGAATGATATACCCGAGTTCTATTTAGTGTGAACAATTCAGCAGGAAAAAGTTGCATCGACCAAGAAAGAACTGCGTTGTTTCAGA  
AAAAGAAAGATGCCGTCACCCCTACGGTAAAGCATTTTAAAGCAGAAACATTCAGATAAACAAGGAATTTCTACACGTAATGAGGCATAAAAA  
AGCTTCAATAACTTCTATGAGCAATAAGCAGTTTCTGCCACATGGCATGTATAGACCCACAAATTTCCAGTGAGTTAAAGCCAAACAGAAAAATG  
CATCATCCCTCAATCGCTCAACATAAATTTGACAATTTAAATACAACAGGCAGTACTGAAAGGAAACATACACAAAAGCTCAGAACTGTCAT  
TGAAATAATCTAATAACCGAGTTCAATTAATGTAATTTATGCAACTGGCCTATGTTCAACATGATATACAAAGGAATTTGAAGCTTATTTCCC  
TCCTCTTTCTTTCAAACACCAATAAATAAATAAAGGCTCTACATGACAAGATAAGCATCGGCTTTCTTCTGCACTGATTAAGTTGTTCT  
CAAAAGTTCCGCGTCTTCTTCTCACACTGCACAAATTTTCACTTCTTTTCCGAAAAAGCCCTTTAGTATGCCCCAAAACCCCTATTAATTAATCT  
TATGCTTTTCTATCACAGAAACAAAGCCACACAGACCAGAATTAACAACAACTCCAATCTCAGCAACCTT

>transcrit\_PtGLIM1b

CAGGACAGTAAGGCAAAAGGCTAAGGACACAACCAAAACAGAAAAATATCACAATCGGTGCACAAGTGTAAAGGTTCTTGTCTTTATCTCTTCAT  
CTTTGAGGGGGGAAAAGAATATGGCAGCCTTTGCAGGAACCCAGCAAAAATGTATGGCTTGTGACAAGACAGTGTATCTTGTGATAAGTTAAC  
GGCTGATAACAGGATTTATCACAAGGCCTGCTTTAGATGCCACCATTGCAGGGGTACCCTTAAGCTAAGCAACTACAGTCTTTTTGAAGGGGTT  
CTGTACTGCAGACCTCACTATGACCAACTCTTCAAGAGAAGCTGGCAGTCTAGACAAGAGTTTCAAGGAACACCAAAAATTTGTGAAACCAGAAA  
AACCTGTTGATAATGAGAACGCAAGAGTCTGAAATTTGTTTGTGGCACCAGAAAAAATGTGTTGGGTGCAACAAGACTGTCTATCTCAT  
CGAGAAGGTTACAGTAAATGGGACTCCATACCATCGGAGCTGTTTTAAGTGTACCCATGGAGGTTGTACTATTAGCCCATCCAACCTACATTGCA  
CATGAGGGAAAACTCTACTGCAAGCATCACCATATCCAACCTTTTCAAGGAGAAGGGAACTACAGCCAGCTTGAATGAACGCGAGAAAAAT  
CAGTCACTGTGAGTATCACCCGCTGTGGAATTTGCCGCTGAATCATAAATCAACTAAAGACTACTGTCCCGCTCTCTAGCCCTTTTCAGTCTCTC  
CTACCGTAAATGGCATCGCAGGCTTTCATGGATTTTCTCTCTCTATATGTCGAAGATTGGTTTTCCGCTTGCAGAAATGGATGTGTTAGTGT

# Annexes

CCCATGGCTGTAATTTTGTGGGAACCTCTGAGTGTGATGTGTCAAGTGAATTTCTGCCAAGATTGAGTAACCTGAGATACTACCTGGCAGATT  
TGATCCCTCTGTTTCTCAAATACGTAAGATTTCGGTATCCGTAATGAT

>CDS\_PtGLIM1b

ATGGCAGCCTTTTCAGGAACCCAGCAAAAATGTATGGCTTGTGACAAGACAGTGTATCTTGTGATAAGTTAACGGCTGATAACAGGATTTATC  
ACAAGGCCTGCTTTAGATGCCACCATTGCGAGGGTACCCTTAAGCTAAGCACTACAGTTCCTTTTGAAGGGTCTGTACTGCAGACCTCACTA  
TGACCAACTCTTCAAGAACTGGCAGTCTAGACAAGAGTTTCGAAGGAACCAAAAAATTTGTAAACCAGAAAAACCTGTTGATAATGAGAAC  
GCAAGCAAAGTCTCGAATTTGTTTGGCTGGCACCAGAGAAAAATGTGTTGGGTGCAACAAGACTGTCTATCCTATCGAGAAGGTTACAGTAAATG  
GGACTCCATACCATCGGAGCTGTTTAAAGTGTACCCATGGAGGTTGTACTATTAGCCCATCCAACCTACATTTGCACATGAGGGAAAACTCTACTG  
CAAGCATCACCATATCCAATTTTCAAGGAGAAAGGGAACCTACAGCCAGCTTGAGAATGAACCGGAGAAAAATCCAGTCACTGTGAGTATCAC  
GCTGTGGAAATTCGCCCTGAATCATAA

>PtGLIM1b

MAAFAGTQQKCMACDKTVYLVDKLTADNRIYHKACFRCHHRCRGLKLSNYSSFEGVLYCRPHYDQLFKRTGSLDKSFEQTPKIVKPEKPVNEN  
AKVSNLFAQVREKCVGNKTVYPIEKVTVVNGTPYHRSCFKCTHGGCTISPSNYIAHEGKLYCKHHHIQLFKEKGNYSQLENEREKPNVTVSIT  
AVEIAAES\*

**PtβLIM1a gène LG\_XII:9413188-9415996**

>PtβLIM1a\_fgenesh4\_pm.C\_LG\_XII000273

TATTATTATTATTTATAAAGAGTTAAGAGAGGCTAGAAAACTATCCTGGATTTTTTTTTTTTTTTTTTTGGATGAGTCACAATTTCTCAATTGGA  
ACTCATGAGCTGTAAAACATGGAAGCAACTCATGAGTCAACAAATAAAATATAAGTTATAAATAATAATAATTTTTTAAAAATTTATAAAGT  
ATGTTAATTTCTCTAAATTTTTTAACTGTGATAAGATTAACAAACAAATCCTAATAAAATTAACAATTAATACTTGTATGTACTTAAAGAG  
GACTAGATCAGAGGATATATATAGTAGTATGATGCATTTTTAATCATGTTTGAATGTCATATAATGTTTTTTATAGTATTTTTATTTAA  
AAATATAAAAAATAAAATTTTAGTTTTTAAATTTATTTTTAATAACAGTATATAAAAACCGTATAAAAATTTAAAAAAATAATTTTTAAAA  
AATCAAAAATTTGAAATGTTTTTTTTTAAACGAAAAACAAATCTCGTGAATTAATTTACTATTTTCTGCCTGAAATAAAGCACAGTACTGTTT  
ACGTCACCTTCCATGATTCAAGTCTAGTAGCTGGACATCTGTGAAAAATAAGGCAATCATGTAGTATAGTCAAACTACTTTTGTGATAAAAT  
ACCAAATTTGCCGCCAACAAAGACTGCTTGCATGTTTTCTAATGGATCTAAAGTGCAGTACTGGGAAATGTTTTTGTATCGTTAGGTTTCTCTC  
TGCTGTCTGAGGCTGAGCTGTAATCTCTAACGCACCCACAGCAATTTTGGCAGCCTTCTCTTTTCATCATCATCCCCCAATCATTCTCTC  
TGCTAGCAGAAAAGCTTATGATATTTGAAGAAGGAGAAGAATCCATCGAAGAACAGGAGACAGGGAGATAGCGCTTGTAGCTAGATAGAT  
ATCATCAATCATTATCTACTATTATCCATCAGCGTACTTAACAGAGTAAAGAATTTACCTATGGCACCATTGCGAGGACAACCCAGAAGTGTA  
AGTCATGTGAGAGACTGTATACCTTTGTGGATCAACTACTGTGATGACAAAGTCTATCACAAGGCTTGTTTTATAGTGCACCCTGCAAGGG  
TACCCTAAAGGTAAGTTACATAGTCCCTTATTTTATAGACACATGAAGAGTCTGCAGAACTCCAAGAGCTGAGCTTTATCATTTGATATATCC  
CTTTTTACTTGTATACATACATATTTATAGCCACCCTCTGTTTCTGATTAATTTTGTAGTAGTGGTTATAGTATTTTGTAGTATTTGATTTGAGGAGGA  
GTTGTTTTAGGTGGATGATTGTGTGCTTGTGTTGGTGTGTCAGATTAGATGCTAACTTGTCAACTAAAAGGTGGTTTTTTTTCTCATTTTTATTT  
CTTAAGCTGATTTGTTTTTCCAACACAATTTATCCAGCATATATTTCTCAGACAATATAAAATGTTTTTTGTTTTTGTCTGGAGTCCAGATT  
GAAATTTGATTTTGCATTTGATTTCACTCCGCTTCTTTTTTAGACCATTTCAACTGGATGGTCTACAGATGGGGTGTGTTAGATCCTTTGATGT  
TAGATGTTGTTTTGTCTTAGTTGTTGATGATAGTCACTTTCTTCGGCCATGAAATTTGCTTGTGGATCCTCGATTTTTGAGTGGTGTATCAGG  
GAGTTCTGCGCCGCCACAGGCTGCCAAGGAGAAGGAGGTTGGGGTGAGAGAGATCTCATTTTTCAAATCGTTGCCAAAACAAAATGAC  
TTGTGTGTTGTTAAAAGCAATTTCTCGGTTGCTTTTTGTTTTGTCGGCCTATTGGGTAGTTGGGAACAGAACAAATGGAATTTTATTCATTCCA  
AACCGTAAAGCTGATTTGAACAGTCTTCCCTGCAGATATCAAATGCAAACTGCATGTTTCTGTAGTCCCTTGGGGCGTGAATCATATATTTATCC  
TATTTGAACCTTTGTTTAACTGGTTTTAACTGAATGATTAAGTGCATGATGACCTGCAGCATGACAGTACAGATACTTATAGTGAATTT  
TTTTCTGTTTTCTCTGCCATAAACTGACCCGCTAGGTACCACATAGAATGTGTTTTAACTGAACAGCGTGAATTTGCAAAATTTTATGGCA  
ATTGCATTTGAATGTGAACACATTTGAAAGGAACTTACTGTCAAAGGATTTCACTCATCTGTCAAAGATATATTGAACTTCTCTTCTATTGGAA  
CGGTATATGGATCGGTTATTGAGTTGCCTATTCTGTTCTAATCTCATGACACTTAGAGTTTCACTCTGCTCAGTCTTCTCAGGTTGAACCT  
GGAGTTGTTCCAGCAATAAAAGACATAGCAACCCGATAGGAAGTAAATTTAGCTAGTATCAGAGAAGGTTGATCAGAGAAGGTTGATGAAACTA  
ACGAATATCAGAAATAGGAATACTGACATCTACCTTAGGATTTCTGTTCAAAATAATCCATGTTAGCCATGTTCCAGAACTTACTATGAAAA  
CCAGAGATGCAAAATACTTTTGTGCATACTATAAAGATTTCAAGAAGCTCGGGTTCCAGGTTCCGAAGAAAACCTTCTGTCTTTTTCTTGTAT  
CACAAAACATGAAGTGTGTTTTCTACGATGTTTTCTTTTTAATTAATCTAAACGCTTTCTTTTTCACTTTGCGAGTTGAGCAACTACTC  
ATCCTTTGAAGGCGTTTTATATTGCAAGCTCCTTCGATCACTTTTAAAGATGACAGGCGCTTGGATAAAAAGCTTCAAAGTGAATACCCG  
TTATAATTTTTCTTTAAGCGCATTTCCCTGCTGTGATCTGCTTGTGTTAAAGTTATTCAAGCTTATTATATTTCCCTTCCCGGAC  
TTCAAAAATGTTAGAGTTGACCGGCTGCGGATCAGTACCCCGAGTCAAAGCTTAAACAGCTGATTAACACTTCATTGCAAGTTGCAGCTA  
TTTCATATTTGATTTGTTATTGTTTGAATGACTATTAAGATGCTTCTTTTTGTTGAATGGAATGCACAGGTTCCAGAGCAACAGCAAAAGTTCA  
AGTATGTTTCCGGAACCTCAGGAGAAATGTGTTGCTGCAAGAAAATGTTTACCCTCGAAAGGTTAAATGGTTGCTTTCTATAATAGTAA  
AATGCAATGATTTGACAGATAAATCTTTTTTTTTTAAATTAAGCAAAATAAATCTGTTCAAATGGTTGACAGAGTGGTTTGTGATTTGTTTATG  
AGCATGTTCAAACTAATGACAGCAGCTAGAAAGCAATGCACCAAGGATGCTAAGAATTCAGAGTTTTTTCTGTTTTATATCAAATGTTT  
TGACAAATCGACATCTTTGCATCAGATGTTCTCCATTTTTTACTCCATCAGTACTCTTTAGAACTTAAAGCTCCAATTCAGAAATCAAAGT  
CCATGTAATCTTACCCCATGATGTTTGTCTGTTTTCTAATCTCATAAGTAACTTATGATGTCATGGATAAAGAAAAATCTTATAGT  
TTCTTGTCTCCAACTTACATCTTTCAGGTTGGCAGTTGATGGAACCTGATCATAAGGCTTGTTCAGATGCGCATGAGGAACTGTTGTCATCA  
GCCCATCAAATCTGAGCCATGAGCATCGTCTGTATTGCGAGCACCATCATAATCAGCTCTTCAAGCAGAAGGAACTTCAAGCACTTGA  
CAAGCAGCAACATGTTACACCAGTGAAGGAGACTGCAGCTGCTGAATGCACTGCATAATTTTCTATAAATCTATGAGTGTGTGACAACCTGAAA  
AGTCTTATTCCAAGAAAAGCTATTATAAAGTGAAGTGAAGAGCTATTATGAATCTGAGCGTATGAAAGAAGGCGAGTGTGTTAGTTGCGG  
AGTCCAGTTTGTGCGCCTACCGTTTTGAGTGAATTTTTGAGTGAATTTTAAAGTGAATTTGAGTGTATATAGTTCTTTTATTTACAACTTAT  
TATTCATGTTGAAGCTTATTTCTATTTCTATGAATGTAGTGTCTGCGGCTAT  
ATTAATATCTTTTTTGTGATTTTTTTTACTTGTGTTAGCTTTTTTGAATACCAATATTTTTTAAATATATATGATGTTAAATTTTTTATGA  
GATTAATGATTTTATTTATTTATATTTATATGAAATTAATTTATTTTTAAAAATAAAACATTTATTTTTTATAATATTTTAAATATG  
AGCAGCTTGCATAGAAGTTTTTTCTTTTTATTTAATCAATAAATTTATGTTATATCGTATATCGATTTTGTGAAATGATTTTCTATAAATAAAT  
CATTAACACAACCGGGTAAATAACATGATTAATAATCTTGAATTTATTTTTTAAATTTTTTGTATTGTTATCAATTTTTTTATTAAT  
ATAAATAATATAGTTTATTTGATAGATATATGAAATTTTTTATTTAAATATATAGACTTTTTTTAGAAAATAATCTACATAACCATTATTTT  
TTTATATAAAAAAACATTTACTACCCGTGAAATAGATAGTATAAAGACACAGTGCCTGGCAACACATGGCAATGAATATAGTGTACGAC  
TAAATTTGCATGACAAAGATGAGAGCTCCATTGCTCTATGGAAGCCAAAGCAACATGCAAAAACCTTAGGAGTTGGGCTAATAGCCCAATAAGCCCA  
CAACAATCAAAAC

>transcrit\_PtβLIM1a

GCTGAGCTGTAATCTCTAACGCACCCACAGCAATTTTGGCAGCCTTCTCTTTTTTCATCATCATCCCCCTCCAATCATTTCTCTGCTAGCAGAAA  
GCTTATTGATTTTGAAGAAGGAGAAGAAGATCCATCAGAAAGACAGGAGACAGGGAGATAGCGCTTGTAGCTAGATAGATATCATCAATCAT

## Annexes

TATCTACTATTATCCATCAGCGTACTTAAACAGAGTAAAGAATTTACCTATGGCACCATTGTCAGGGACAACCCAGAAGTGAAGTCAATGTGAGA  
AGACTGTATACTTTGGGATCACTCACTGCTGATGACAAAGTCTATCACAAGGCTTGTTTAGATGCCACCCTGCAAGGGTACCCTAAAGTT  
GACAACTACTACTCCTTTGAAGCGTTTTATATTGCAAGCCTCACTTGCATCAACTTTTAAAGATGACAGGCAGCTTGGATAAAAAGCTTCAA  
GGGACTTCAAAAAGTGTAGAGTTGACCGGTCTGCCGATCAGGTCCAGAGCAACAGCAAAGTTTCAAGTATGTTTGGCGAACTCAGGAGAAAT  
GTGTTGCTTGCAGAAAAGTGTACCAATCGAAAGGGTGGCAGTTGATGGAACATCGTATCATAAAGGCTTGTTCAGATGCGCTCATGGAG  
CTGTGTCATCAGCCCATCAAATTCGTAGCCATGAGCATCGTCTGTATTGCAGGCACCATCATAATCAGCTCTTCAAGCAGAAGGGAACTTC  
AGCCAACTTGACAAGCAGCAACATGTTACACCAGTGAAGGAGACTGCAGCTGCTGAATGACACTGCATAATTTTCTATAAATCTATGAGTGTGT  
GACAACTGAAAAAGTCTTATTTCCAAGAAAAGCTATTATAAAGTGAGTGAAGAGCTATTATGAACCTGTAGCGTGATATGAAAGAAGGCAGTG  
TTAGTTGCGGAGTTCCAGCTTGTGCGCGTACCGTTTTGAGTGAGATTTTGAAGTGTTAATAAGTTGAGATGATTTATAGTTCTTTTATTTA  
CAAGCTACTTATATTCATGTTGAAGCTTATTTCTATTTCTATGAATGTAGTGTCTTGGCGCTATATATATATATATATATATATATCTTGCA  
TTTATATTTTGATT

>CDS\_PtβLIM1a

ATGGCACCATTGTCAGGGACAACCCAGAAGTGAAGTCAATGTGAGAAGACTGTATACTTTGGGATCACTCACTGCTGATGACAAAGTCTATC  
ACAAGGCTTGTTTAGATGCCACCCTGCAAGGGTACCCTAAAGTTGAGCAACTACTCATCCTTTGAAGCGTTTTATATTGCAAGCCTCACTT  
CGATCAACTTTTAAAGATGACAGGCAGCTTGGATAAAAAGCTTCAAAGGGACTTCAAAAAGTGTAGAGTTGACCGGTCTGCCGATCAGGTCCAG  
AGCAACAGCAAAGTTTCAAGTATGTTTGGCGAACTCAGGAGAAATGTTTGCCTGCAAGAAAAGTGTACCAATCGAAAGGGTGGCAGTTG  
ATGGAACATCGTATCATAAAGGCTTGTTCAGATGCGCTCATGGAGGCTGTGTCATCAGCCCATCAAATTCGTAGCCCATGAGCATCGTCTGTA  
TTGCAGGCACCATCATAATCAGCTCTTCAAGCAGAAGGAAACTTCAGCCAACTTGACAGCACGAACATGTTACACCAGTGAAGGAGACTGCA  
GCTGCTGAATGA

>PtβLIM1a

MAFFAGTTQKCKSCEKTVYFVDQLTADDKVYHKACFRCHCKGLKLSNYSSEFVLYCKPHFDQLKMTGSLDKSFKGTSKTVRDRSADQVQ  
SNSKVSMSFAGTQEKCVACKKTVYPIERVAVDGTSYHKACFRCAHGGCVISPSNFVAHEHRLYCRHHNQLFKQKGNFSQLDKHEHVTPVKETA  
AAE\*

**PtβLIM1b gène LG\_XV:5205959-5209047**

>PtβLIM1b\_estExt\_fgenesh4\_pm.C\_LG\_XV0182

AATGTCGCCACAATTCAAGAAAAGCTCGTGCAAGAACTCCACTCTCCCTCGTTTTCTTTGCGAAAAGCAGACTTTATTTTTGGATCATTAAAGGGAG  
GCAACATATATCCACTACGAGTTTAAAGGGACTAATTAAGTAGTCAAATCCAAATTCAAAACAAGAAAAATATAATGATTTGTGCATTATTAAGAG  
ATAACAGTAAGAATTAATAAGATAAAATCTTTAATGTTAATAAGAATTAATATATATATATATTTTTGAAGGAAAGTGTGTTTGTCAAAT  
AACATGCGAACCGACTTAAACTTAAATTAGTATTGGTATGTTCTTTTGAATAATTCAAAATAAAAAGAAAGCTGGTTCAAAGAACACCTTAC  
CATGATAAGTTGGTTGGTAAATTTTTTTTTTTTTTAAATTAATGTTTTTAAATGTTTTTGTATCGATTTAATGTTAATAATTAATAATTA  
AATAAAATTTTTAAAAATAAAAATATTTAATATATTTTTTAAATAAAAAATATTAATGCTGAACAATCTCTATGAATAAAAAGTATAGTGCT  
AATCTTATAGGTTGATGTTGAACAAAGTTTTGAAAAATTTTTAAATATTTTTTGTCTTAAAATATTTTTAAATTAATTAATATATTAATATTA  
AAAATAAAATTTTTAAAAGTAAAAAATATTTCAATGCCCTGTAAGTAAAAACAGGATATGTTGCAATCCATCTAGTGTGAACATAGATTTAGA  
TTCTTGTAAAGATAAAATCACAGAGCTTCTTACAGTCTATGACTCAAGTCTAGTAGCTGACACCTGCGAAACGTAAGGCAATAATGTAAT  
ATAGTCAGCTACTTTGGTTTTAGAAAATGACCAAAATGCCCAACAAAGACGGCTTGCATTTCTATTGCGTCTACGGTCTACTTGTAAATTA  
GTAATACGTATCTGATGTTAGGTTTCTCTCTTTTTAACTCTCTGGGTGTGTAATCGCGAACGCACACAGATTTTGAAGTCTCTCTCTCT  
TACCTTTTATCACCACCACCAATCTTTCGAAGTTTTTGAAGAAGAAGAAGATCCATCAGAAAGATAGATAGATAGCTAGCTAGCTAGATGATA  
TCATCTTCAATATCTATATCACCCTTTATTTGACAAAAGTAATCATCGATGGCAACATTTGTCAGGGACACTCAGAAGTGAAGCAT  
GTGATAAGACTGTGACTTGGTGGATCACTACCGTTGATAAAGTAATCTATCACAAGGCTTGTTTTAGATGCCACCCTGCAAGGTTACCCCT  
CAAGGTTGATACCATCTCCCTTTGTGTGTGTGTGTGTGTGTGTGTATCACAACATACATGCAGAAGTCTGCTCAAAGATACAGATATAATAT  
TGACATCTCCTCTTTTACTGGATCATACATATTTATAGTCACTTTTACTTCAATTTCTTCACTTTTGTAGTATATTTATAGAGAAGGTTGT  
GTAACCTGATGATGTGTGCT  
GCTTTCCAAACATGATTTCCCAACATATATGGCAGATAGTACTCGAACAATAAGAAATTTTTTTTTTTTTTTTTTTTTTTTTTGAAGTGTGAT  
GATCATGATTTTGCATATGATTTCACTTTTCTTTTTTCAATTTGGATGATCTATAGATGGGGTTTTTTAGACTCTTTGATATTAATGCTGTT  
TGCTCAGTTGGTTGATGAGTGGTCCCTTTTTTGGCCATGAAATTTAATTTGTAGATCCTCAGTTTTTGGAGCGGTGATGGGTATCCCCACAGGT  
TGTCAGGGAAGGAGGTTGGGTTGAGAGAGGTCACATTTTCAAATCTTTGTCAAGCAACAATGACTGTGGATGGGTTAAAAGCAAT  
CTTCAATTTGTTTTGTTTTCTGTGCT  
TTTCTCAGATATCAAATTTAAAACAGCAAAATGGAATTTTGCATATACAGAGTGTCTATTTTGTAGTAAATTTTATGGCAATATTACATCA  
TAATATGGAGAATCTTATAGATTACCAGGGAGAGATACATGAGTTATTTGTCATGTTACAGGGACACAGCAGGACCTTAGAAGGTTGAGTAA  
ATGCAAAATGCTGTTCTTTAGTCTTGGACCGTGTGAATCATATTTGTTTCAATTTGAACTTTGTTTCTGTTTAACTGAATGATTTAAGT  
TGATATATATGCT  
GTACCACGAAACATGGTTAACTAAATCAGCATGCAATTTGCAAAATTTTATGGCAATTTGAAGTGGGTGTGAATACATAGGGAAGAAAACCTTAC  
TGTAAGGATTTTACCTCATCAATCAAGATACATTTGAACTTATCTTTATTTGGAAGCATATATGGATCATAATGAGTTGCGTATTCATTTCTCC  
AATCTCATGACTTTCTGAGTTACGTTTCTGCTTGTACCTCTGGGGGTGAACCTGGGGCTTGTCCCGAAATCAAAGACATAGTATAGCCATA  
TATGAAGTAAATGAGTGTGAAAAAAGGTAATCAGAGATTGAACTACCTGTTGTAACATCCGCTTGTCAATTTCAATTCACGGGCTATATTTTCTCTCC  
CTGGTCAAACATAGTGCATAGTAGCCATATGCAGAAACCTATTATAAAGATTTCCGGCAGCTTCCGGATAGGTTCCCTAAGAAAACATGGTGCCC  
TTTTTCAATTTGAAAACAGATATGAACTCGTGTTTTCTATGATGTTTTTCTTAAATTTGATCTAAACATGTTCTTTTCACTTTTGCAGTTGAGC  
AACTATTCATCTTTGAAGGTTTTATATTGCCAGCCTCACTTTGATCAACTTTTCAAGATGACAGGCAGCTTGGATAAAAAGTTTTGAAGGTG  
AATACCAGTTATTAATTTTCTGTAAGGCTGGAAGGCTAAGGATTAAGGATAAATAAGCAAAATTAATCTGTTTCAAGATCATTGAGTGGTTGCT  
TTGCGAGCATGTTTAAAACATGAAATAGAAAGCTATGCTACCAAGGATGTCGAAGAACTTTAAGTTTTGGGCTGGAACACAAGTCATATATTT  
TTTCTATTTTATCTCAAATGTTTTTCAAAATGACACTTTGATCGCTTAGTCTCCATTAATGCTCCAATTCAGTCTGTGAGATCTCTAC  
CCATGTGTTTGTCTAAGATTTCTCAAATCTCAAGCTCAAGTGAATGTTGATGTTGTTGAAAAAGAAAAATAGGAACTATTTCTATAGTTTCTT  
GTCCCAACTTTTGTGCTATTTCAAGTGGCAGTTGATGGGACATCTTATCATAAAGCTTGTTCAGATGCGCTCATGGAGGCTGTGTCATCAGC  
CCCTCAAACATGATAGCCATGAGCATCGTCTGTATTGCAGGCACCATCATAATCAGCTCTTCAAGGAGAAGGGAACTTCAAGCAATTCGGCA  
AACATGAACATCTTACACCAGTGGATGAGACTGCAACTGCTGAATGATGACTGTATAATTTTCTATAAATTTTGTGTGATAAGTACTCAGAGAG  
CTTATTTCCATGTAGAGCTGGATGCAATTTGCTGAAGTGCATGAAAGCAATAAAGCTGTAGTATGTTTGAAGAAAACGGTATTCATTT  
CGGAGTTCAGCTTGTGCTGTAACATTTTGGAGTGTGAGTTTCAACTTATAAATAAGCTGAGATGTTGTTGATGCTTTTATCCACAACAT  
AATTTTTCTGTATTTATTTTCAAGCTGAAGCTTATTTATATGATGTCAGTGTCTTGTGTAACCTTGTCTTTTATCTGGAGTGTCTGTGCTGCT

# Annexes

GGCTTGTTCAAATTCCTTATACAATTCACATATCTAGACCGTTAAAAGCTATCCAATGTTCCGAATCCTTAGATTTCATGTTTCATAACAATTAATC  
AAGCTCCTTTTCATTTTTCATAACAATTAATAATAGATTTGGATCATGATTACAAAATCAGTCAGGGCGCAATACCTAACGAATTACGTTGAAA  
TTATTAAGCCCTTAGTTTCATGACTGTCGTCGAATAATTTCTACTGCTCTGCAAAAAGATAAAAAGGTCATGCTAAGGCTAATAACTTGCAGC  
AACTTAATTAACAAAACAGGCTGTAGTTAAATCAGTGTGCCACTCATAAAACATGTTTTACAGTGTGTGTGCCCTTTTTTTTTTTTTTTT  
AATTTCAACATTAATTTAGAAAATGTGATTTGTTATTTTATTTTGGTGTATATTTGGTATTAATTTTATTTTAT  
TTCAACTTTCAACATTAATTTACTAGAGGTGGGGTTTCATATATTTTTTTTATATATATAAAAATATCTAATCTCAAAGCTTAAGTCAC  
AGGTTTAAACAAATTAACCTGGTTGACCCGATTTATTTTGTAGACCAATTTTTTAATTGATTTTTTTCTGATTTCATCATTATTTATTTGGG  
TTGGTTAGAACT

>transcrit\_PtβLIM1b

CGAACGCACACAGATTTTGGAACTCCTCTCCTCTTACCTTTTTATCACCACCACCAATCATTTCGAAGGTTTTTGGAGAAGAAGAAGATCCAT  
CAGAAAGATAGATAGATAGATAGCTAGCTAGATATCATCTATCAATATCTATATCACCCGTTTATTTGACAAAGTAAAGTAATCATCGATGGCA  
ACATTTGACAGGGACAACCTCAGAAGTGTAAAGCATGTGATAAGACTGTGTACTTGGTGGATCAACTCACCGTTGATAACAAATTTCTATCACAAGG  
CTTGTTTTAGATGCCACCCTGCAAGGGTACCCTCAAGTTGAGCAACTATTCATCCTTTGAAGGTGTTTTATATTGCCAGCCTCACTTTGATCA  
ACTTTTCAAGTTGACAGGCAGCTTGGATAAAAAGTTTTGAAGGCCTCCGAAAACCTGTTAGAGGTGACCCGCTGCGCATCAGGTCCTTAGCAAC  
AGCAAAGTTTCAAGTATGTTTGTGGAAGTCAAGAAAATGTGTTGCTTGAATAAAAACCTGTTTACCCTCGAAAAGGTGGCAGTTGATGGGA  
CATCTTATCATAAAGCTTGTTCAGATGCGCTCATGGAGGCTGTGTATCAGCCCTCCAACATGTAGCCCATGAGCATCGTCTGTATTGCAG  
GCACCATCATAATCAGCTCTTCAAGGAGAAGGGAACTTCAGCCAATTCGGCAAAACATGAACATCTTACACCAGTGGATGAGACTGCAACTGTCT  
GAATGATACTGTATAATTTTCTATAAACTATTGTGTGATAAGTACTCAGAGAGCTTTTATTTCCATGTAGAGTCGGATGCAATTTTGTGAAGTGC  
ATGAAAAGCAATAACAAGCTTGTAGTATGTTTTGAAGAAAACGGTATTCATTTGCGGAGTTCAGCTTGTGTGTAACATTTTGTAGTGTGAGT  
TTCTAACTTATAATAAGCTGAGATGTGTGTAGTCTTTTTATCCACAAACTATAATTTTTCTGTATTTATTTCAAGCTGAAGCTTATTTATAT  
GGATGCAGTGTCTTGTGTTG

>CDS\_PtβLIM1b

ATGGCAACATTTGACAGGACAACCTCAGAAGTGTAAAGCATGTGATAAGACTGTGTACTTGGTGGATCAACTCACCGTTGATAACAAATTTCTATC  
ACAAGGCTTGTTTTAGATGCCACCCTGCAAGGGTACCCTCAAGTTGAGCAACTATTCATCCTTTGAAGGTGTTTTATATTGCCAGCCTCACTT  
TGATCAACTTTTCAAGTATGACAGGCAGCTTGGATAAAAAGTTTTGAAGGCCTCCGAAAACCTGTTAGAGGTGACCCGCTGCGCATCAGGTCCTT  
AGCAACAGCAAAGTTTCAAGTATGTTTGTGGAAGTCAAGAAAATGTGTTGCTTGAATAAAAACCTGTTTACCCTCGAAAAGGTGGCAGTTG  
ATGGGACATCTTATCATAAAGCTTGTTCAGATGCGCTCATGGAGGCTGTGTATCAGCCCTCCAACATGTAGCCCATGAGCATCGTCTGTAT  
TTGACAGCACCATCATAATCAGCTCTTCAAGGAGAAGGGAACTTCAGCCAATTCGGCAAAACATGAACATCTTACACCAGTGGATGAGACTGCA  
ACTGCTGAATGA

>PtβLIM1b

MATFAGTTQKCKACDKTVYLVLDQLTVDNKFYHKACFRCHHCKGLKLSNYSSFEGLVLYCQPHFDQLFKMTGSLDKSFEQTPKTVRGRSADQVL  
SNSKVSMMFAGTQEKVCACNKTVYPLEKVAVDGTSYHKACFRCAHGGCVISPSNYVAHEHRLYCRHHHNQLFKEKGNFSQFGKHEHLT PVDATA  
TAE\*

## PtWLM2a gène LG\_VIII:3623285-3626671

>PtWLM2a\_estExt\_fgenesh4\_pm.C\_LG\_VIII0239

GATACTATTGCAATCAATGACAAATATTTTCGATATTTAACTTATGTTTTAATTTTATACATATTTATTTTTATTTATATAAAAATTAATTAATAAAA  
ATTAACATCATCAAAAATATTTAGGATGCCAACTCATCTCATTCGTTATTTTCATATGAATAATGAAGGAGGCAATCTCAATTTTTCTTTGCATC  
CCCCATCTTTTTTTTTTCTTCCAGTCTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAACATTTATTTTAGTAAATTTTTAGTTTTACTTTTTTCATGAAAT  
GTCGTAGTCCCAAATAAAAATTTCAACATTTGGTTTACTTGTATGTTACAAGTGTATTTTTTTTTTTTTATATAATAAAGAAAAAAATTAAT  
TTAAAAAACAATAATTTTAAACCTAGTATAGTTTATTTACCGAACAGTGAAATTTAACAGATTAATTAATAAAGTAAACTAAATAAAT  
TACATCTTAATAATAAAAATAAATTAATTTGATTTTTTAAATCAACCATATTTTTTATTAATTTATTTGATTTTTTATCTTCTTAAT  
AACTGTATGAGGTCAAGGTAAATTCACACGTAGATTTGATTTTTTAAATATTTTCCCAACCATCGCCCTAATTTTTGAAAGTCTTCCACAATAAG  
TCTCTCAATGGAGTCCCTCGTTTAAATTAATATCAGTATATCACCAAATCTCAGCCTCCATTTAATTTATGTTTGAATTTAGTTTGAAT  
CTGGGAAATTAGTAAACATCAATTTAGATGAAGAATTTGGGAAACCTTTTGTGAGGAGAATTTGATCAAGGATAGGGTAAATCGTTTGGGGACATAT  
TTGCTCTTTCCATTTATTTTAAACCACGCGCAGCATTATCAGGGTAACTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA  
CCAAAAAAGATTTAGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA  
GCTTCCAGCGAAGAAGATATTTTCATCTCTCTTACTCTTTCTGTCTCTCTAAATTTTTTATAACCTCTGAAATCTCTCTCCCTCTATCAT  
CATCATCATCATCCACTTTGTTGATCCGCTCCGCTACAAAAGCGCCCGCCCAATCCCCGCTCCCTTCACCTTTCTCTCGACCTTTTCACTTT  
CATTCAATTTCTCTCCACTCACCTCCCTTGACCCTCAGATCGAAACAATCCCCTCTCTTCTTAACTGCAATTTGGTACGCGCCCTCTCTT  
TGATCCTTTGGGAAAACGAGTATTTGACTGCATTTGACTGCTTCTGTTTTCTGTTTTCTGTTTTCTGTTTTCTGTTTTCTGTTTTCTGTTTTCTGTTTT  
TTATTGAAGTCTTGCATTTCTTGAAGTCTTAGATCGGTTACCAATCTGAAACCTGAAGTTTACATTCATTTATATAATTTCTGATCTAT  
TGTGCTCTTTTGAAGTTTAACTTTTATTTTCCATGAGTCTGACTGATTTCTATTAATCTGCAAGGTTCAATCTTTTTTATCCATTTGATTT  
GATACTTAGTTGATCTGTTGGTTTGTAGACTGGCAGGTGATCTGAGCAGGTAAACAGTACACAGTCAATCAGATCTAAAGGGTAAACATACATAT  
CGCTCGAAACGATGTCGTTTACTGGTACCCAACAGAAATGCAAGACTTGTGAGAAGACTGTGTATCCCATGAACTTCTATCACAAGATGGGGT  
TGCTTACCATAAATCTTGGCTCAAGTGTTTTCACTGCAAGGCAGCTTAAAGTGTCTTATACCGTCTTAAATTTTGGCGTACCTATTTTCTT  
TCCTGTTCTGTTAATCCAGTCAAGAATGTAGCTGGCAATGCTTAAATTTGCAACTGTTAGGGTTGTCGTTCACTTGTGTAGACATGGCATCAAAG  
ACAAATTTACCAAGTGGTTCTGCTTTAGCTGTTAGGTAATGGATTGTTTTAACAATTTGACCTGTTAAATATGTTTTCTCAATTTGGAGTGTAT  
GATAAGCAGCATTTGTTAGGACTTTTTCATATAAAAAACAGTATGATATTTAGAGAGTGAATTTAGGGATAATCGCTCGTGTATCAACAGATGAT  
TTATATGATTTTTCCCGATTTTATTAGAGGAAATATCCGCAGTATGTTCCAGATTTTAAATGGATGGCTAGTAGATTTATGTTTTGGTTT  
GTGCTGATTTAGTTGTCAATTTATGTTTTTCCAGCTAAGCAATTTATTCATCAATGGAGGGTGTATTGTACTGTAAAGCTCACTTTGACAGCT  
GTTTAAAGGAGACCGGTAACCTCAACAAGAATTTCACTGCGCTATGCCCTCATCTCCTGCTGGGATCAATTTCTTTTATTTGATATCCATATAT  
TTTTTTTATTTGTTTTTAAATTTTCAATTAATTTGTTTTTCAACAGCTGCAAAAGTCAAGTGTGAGAAGTTAACTCTGAACTGCAAAATTTCCCTG  
CAAAGACTGTTTGTATTTGGCATTTATTTACTTTTGTGATTTTCAATTTATTTGACTGATGTTGCTGGTGTCTCTGATCCCTATTTGATG  
TTCTGTTTTAGCAAGGTCACCTAGCAAAAGCTGCTGGTATGTTTTTCAAGGACACAAGAGAAGTGTGCTACTTGGGGTAAACTGCTTATCCGC  
TTGAGAAGGTAGGAGTCAGATATTTGTGATATACAGGCATATAAGGTTGATGATAGAACTCCTATCATTAATTTGATTTGCTTGGCTCGGTT  
GAGTTTTTCTCTGGTCGAATGAGGCTGCTCGTTAATCTTGTATTTGATTTAGGACCATATAACAAATTAACCTGAAAGAGAAGTATTTCTACAAG  
AAGTTGGGGCTTATTTGACAAGTCAAAGCGGTGGGGCTACAGATTTCTGCAGTGTGGGAATGCGCAGTGAAGCTGCTTTTTCAAAAATTT  
GAAATTTTCTTTTTTTTTTGTATAAATATTTTTTTATTTTATTTTATGTTTTTTCAGATGTTTTTGTACTGATTAATAAATAAATTTTAAAA  
AATAAATAAATAATTTTAACTTTTCTGCAAAAACACATTTGAACTTCTACCGCTATCACACACTACAACATCTTTGATAGTATA  
GGAATGAAGATTCATGTAACCTTTAATACATGACTTAAGCCAGGGCTATTTTTATCCCAATAAATTTGAACTTGTGATTTCAAGGAGGTTTTCT  
TTTCAGCTTCTTCTTCTTCTTCTTCTTCTTGTGGACATGGATTGAATAGCTTGTGATGACAACAGCATTGATAGAATACATCTCCACATGTTGAAAG

# Annexes

GCTGTGCATCTTAATTGACTATTTAAGTTGTTAGCATCCAATAAATCACAACTCCTAACCCCTAACCCCTACTACTAGCTTGGATAGCAGGA  
TGCTTATAGTTTTCCCTTTCCCTTTTTTTTTTCCCTGTGAGAAAGTGCATAGTTCTTCTGGTAGATGGTTTGATAGTATGATGCTATTT  
CTACCAGTGGTTGTATCATATATCGATATACCTCCCAATCTAAGTTTAAAGAACTAGGCCATTAGTATGCAATATTTGGTTAAATTTAAAGTTTA  
GGCCTCTCAACTACCCTTCTTTGCATACATTTATTTGCCAATTTGGTTAAATTTAAAGTTCCATGAAAATAAATTCATCTCTCTTTGCGTGGA  
AGTTATTTATTTTCCTTTTAGCGAGTTCTATCTGTCTATGCTAAAAAATACCAGGGATCATGGTAATGGTCTGTTCGCCAAAAACTACAG  
GTAACAGTGGAGAGCCAAAGCCTATCACAACTGTTTCAAGTGTCTCACGGTGGCTGTGCTATAACACCATCAAATATGCAGCGCTTGAGG  
GTGATTTATTTGCAAACTCACTTTTCCCAACTTTTCAAGGAAAAGGGCAGCTACAACCATCTGATCAAGTGGCGATCAATGAAGCGAGCAGC  
TGCTCTGTGCCGAGGCTTAAACATCTCTCCCTTAATTTGATATGTACTCGTGTGGCTTTATCTACTGTTTCAGTGTGTTTTCCCTCCAAACCC  
TTCGATTTCCATGGAGGGGTAGGCAAAACCTGGTTCGCTTCTTTTTTAAAAAACTTCTCCCTCAGCTCTCGGCAGGATGTTTTGCTGT  
GTTTGATGATAGATGGGTTTCAAACCTCCAAAATTTGAAAATGACTGGAGGAATTTCAATGGCAATATTTGGATGCTATTTAGTGTCTTCTACT  
TGTGTATCTGGTTTTGTACCACATATTACGAGGAAAATTTGGTGCATTAATATGATATGCGTCTACTCTCTGAACTCATCGTCTCCCTCAT  
TCTTCTATTCTCAATCATGATTGACGCGTGGATTAACCTGAAGGGTTTTCCATGGTGGTGGCAGCAATCATAAACACGAGATTAGTAATTTG  
CTCGACCCAAAAAGAAAGAAAGAAATTAACCACAAGGGGTATGGTCTCACTAGTTTTTTTTTAAATGTTTTTAATGATAATATTAATAA  
TAAAAAATATTTAATATATTTCAAATAAAAAATACTTTTTAAAAATAATTTACTATATTTTTAAATATTTTTATATAATTTCTTT  
TTCAGATTTACTTTAAAAAATAATTTACTGCATTTCAAATGTTTTCAAATGTTTTTAAATAATTTTTTCAAATTTAACTCATTCGAAATC  
ATTTTATTTACTTTTATATAAATAAATAAATAAATAAATAAAGAAATCACATTTTGAATTTCTGAAGTTATTTAGATTTGGATTTTGAAAA  
AAACATTCATATAAGAATATCTGTACAGAGATAATTTATAAATAATCTGTATAAATTTACTTTTTAAAGATTTGTTAAACAAATTTTGT  
ATTGAATGCACAAAATACTGTTTATATACTTCAATGTCAAATAGAAAATTTGAAAAAAGAAATTTGTTCCGGATGAAACTTGAAGATG  
CATATCTGTTAAAAAGAAAAAGAAAAAATCTTTTCCCTGGCAAAATCTGTGTCATTTACCAAAAACGACTCTTTCCCACTTTTCCCACT  
CTGTGATAGCCTGGCTATATAAGAAAGATTCAAATATGAAAAATAAAGACAGAAAGCGGACTGGAGAGGCAAAAAATGTGGAATTCACTCT  
AGCTCTCCATGCCATCTAATTCAGTAAGCTCTCCGTATTTGCAACCCTAAAAATTTTTCAATCTCTAATTTATTTGTT

>transcrit\_PtW LIM2a

AGACAACAAAAGTGTGTTGGAGCCTGAAGTGCTCCAGCGAAGAAGATTTTTTCATCTCTCTTACTCTTTCTGTCTTCTAATATTTT  
TTATAACCTCTGAAATCTCTCTCCCTCTATCATCATCATCCACTTTGTTGCATCCGCTCCGCTACAAAAGCGCCCGGCCAATCC  
CCGCTCCCTTCACTTTCCCTCGACCCCTTCACTTTCACTTTCAATTTCTCTCCACTCACCTCCCTTGACCCCTCAGATCGAACAATCCCATCTCTCT  
TATTCTAATCGCAATTTGACTGGCAGGTGATCTGAGACAGGTAACAGTACACAGTCAACAGTCAACAGTCAAGGGTATAATACATATCGCTCGAAGC  
GATGTCTGTACTGGTACCCAACAGAAATGCAAGGCTTGTGAGAAGACTGTGTATCCCATGGAATCTTATCAACAGATGGGGTGTCTTACCAT  
AAATCTTGTCTCAAGTGTTTTCACTGCAAGGCAGCTTAAAGCTAAGCAATTATTCATCAATGGAGGGTGTATTGTACTGTAAGCCTCACTTT  
ACCAGCTGTTTAAAGGAGCCGTAACCTCAACAAGAACTTTCAGTGCCTGCAAGTCAAGTGAAGTTAACTCCTGAACTGCAAGGTCACC  
TAGCAAGCTGCTGGTATGTTTTCAAGGACACAAGAGAAGTGTGCTACTTGGCGTAAAACTGTTTATCCGCTTGAGAAGGTAACAGTGGAGGC  
CAAGCCTATCACAACTGTTTCAAGTGTCTCACGGTGGCTGTGCTATAACACCATCAAATATGCAGCGCTTGAGGGTGTATTATATGCA  
AACATCACTTTTCCCAACTTTTCAAGGAAAAGGGCAGCTACAACCATCTGATCAAGTGGCGATCAATGAAGCGAGCAGCTGCTCCTGTGCCGGA  
GGCTTAAACATCTCTCCCTTAATTTGATATGTACTCGTGTGGCTTTATCTACTGTTTCAGTGTGTTTTCCCTCCAAACCTTTCGATTTCCATTGG  
AAGGGGTAGGCCAATTTGGTTCTGCCTTTTTTTAAAAAACTTCTCCCTCAGCTCTCGGCAGGATGTTTTGTCTGTTTTGATTTGATAGAT  
GGGTTTTCAAACCTCCAAAATTTGAAAATGACTGGAGGAATTTCAATGGCAATATTTGGATGCTATTTAGTGTCTTCTACTTGTGTATCTGGTTTT  
GTTACCACATATTACGAGGAAAATTTGTGCATAATATGATATGCGCT

>CDS\_PtW LIM2a

ATGTCGTTTACTGGTACCCAACAGAAATGCAAGGCTTGTGAGAAGACTGTGTATCCCATGGAATCTTATCAACAGATGGGGTGTCTTACCATA  
AATCTTGGCTCAAGTGTTTTCACTGCAAGGCAGCTTAAAGCTAAGCAATTATTCATCAATGGAGGGTGTATTGACTGTAAGCCTCACTTTGA  
CCAGCTGTTTAAAGGAGCCGTAACCTCAACAAGAACTTTCAGTGCCTGCAAGTCAAGTGAAGTTAACTCCTGAACTGACAAGGTCACCT  
AGCAAAGCTGCTGGTATGTTTTCAAGGACACAAGAGAAGTGTGCTACTTGGCGTAAAACTGTTTATCCGCTTGAGAAGGTAACAGTGGAGGC  
AAGCCTATCACAACTGTTTCAAGTGTCTCACGGTGGCTGTGCTATAACACCATCAAATATGCAGCGCTTGAGGGTGTATTATATGCAA  
ACATCACTTTTCCCAACTTTTCAAGGAAAAGGGCAGCTACAACCATCTGATCAAGTGGCGATCAATGAAGCGAGCAGCTGCTCCTGTGCCGGA  
GCTTAA

>PtW LIM2a

MSFTGTQQKCKACEKTVYPMELSLTDGVAYHKSCFKCFHKGLKLSNYSSMEGVLYCKPHFDQLFKETGNFNKNFQSPAKSAEKLTPELTRSP  
SKAAGMFSGTQEKATCGKTAYPLEKVTVESQAYHKSCFKSHGGCAITPSNYAALEGVLYCKHHFSQLFKEKGSYNHLIKCASMKRAAFAVPE  
A\*

**PtW LIM2b gène LG\_X:17210641-17213976**

>PtW LIM2b\_estExt\_fgenesh4\_kg.C\_LG\_X0064

GTTTTGGTTCGGTTTTTTTCAGTTTTAGTCTTATAAAACCAAACGAACCGGTTAATTTTTTCAAATTTTAAATGGTTTAAATCGGTTTTTTTCA  
CAGTTCGGTTTTTTTCCAATTTAATTTGGTTTTTTAGTTTTTTTTTTTCCACTAGTTTTTGAACCGTGTGTAATACATATGCTTTATTTGAACACA  
TAGTTAACGATACTATTGAATAATCTTCAATTTGAATCCCTCATTTAATTTAAATATTTGCCCTCAGATATTTAAATATGGTCCGAAATTTAT  
TTCAAATTTGGGTAATTTAGGTAACATGAAATTTAGATGGAGACTTAAATTTGCAAAATATTTGTTGTCAGGAGACACAATCGAAGAACAGCAATTAAT  
ATTAATTTGGAATTTTCAACATTTTTAACACGCGCAGCGTTACCAGGTAATAAATAAAAAAAGAGGGTCAACGTTGTAATTTTTAT  
ATCACAAAAAGATGAGAAAAAAGAAACCCATGATGGAGTTCGTTGAATCAGACAAAAAGTGTCTGCGGAGCCCGTGTGCAAGGTT  
CTTCCAGCGAAGTAGGATTTTCACTCAATCTTTTCTATCAGTCTATGTAATTTTTTGAATAATCTCCGAATTTATTTCTCTTTACATAT  
ATAAATGATTTAGAAATGAAATACTAGTTGTTTTTAGATATATTAATAATTTTGTGATATTAATATTAATAATATACACAAAAAT  
ATAAAAAATAATTTAATTTAAAAAACCACAAAAAGTTTATTTCTCTTTACATATATAAATGATTTAGAAATTTAGAAATTTGAAATAGTAG  
TTGTTTTTAAAGATATAATAAATAATTTTTTATTTTTTAAATTTATTTTGTATTTAATATATAATATAATAATAATAAATAA  
TCCAAAAATTTCTAATAATTTTTTAAACAAAAACAGCCCTTCCGCTGCATCTCTCGCTGCAAAAGCGCCCAATCCCGCTCCCTTC  
GCTAGCCCTTTTATCTTTAAACACCCCTTTCACTTTTCACTTTTTCGCTTTCTTTTCACTCTCGATTTCACTTTGATTTCCCGATC  
GAAATATCCATCTTTTCCGATTTCAATCTCAATTCGTTACGCGCTTCTCTTTGATTTTTTGGCCAACAGATTTCACTTCATTTCAATTTGCT  
GTTTTTTTAAATGCTAATCTTGTCTATTTCTCTGCTAATTTAGCTCATTTTTTAAAAATTTTAAATTTCTTGTGATTTCTTTGAAGTTTGAAGT  
CTTAGATCGGTTACCAATCTTGAATTTGAAATCTTGAAGTTTATTTATTTATAAACACTTCTCTGATCAACCGTGTCTTCAATTTGAAAGTTCA  
ACCTTTATCTTCAAGTCTTAACTGATACTGATTAATCTGCAAGCTTCCATTTGATTGATACATAGCTGATTTGGGGTTTTGGATGGCA  
GGTTATCCAAAGGAGTTAACAGGACAGTCAACAGTCTAGAGGATAGTTTATATAGCTAGAAACGATGTCGTTTACTGGTACCCAACAGAAA  
TGCAAGGCTTGTGAGAAGACTGTGTATCCTATGGAATCTTATCAGCAGATGGGGTGTCTTACCATAAACTTGTCTCAAGTGTTTTCACTGCA  
AAGGCAGCTTAAAGTGTCTTAACTTTTAACTCTGACTTACCTTATTTTCACAATTTGTTTTGTTTAAATATAGTCAAGAATGATTTGGCA  
TGCTTATATCACAACTGTTAGAAATTTGTTCAATTCAGATAAATTTGACCAAGTGTATTCGTTTTAGCTGCTGATTTGTTTAAACACATC  
GACCTGTTAAAGATCTTTGCCTTTAACTAACTGGAGGATGATAAGCAGAAATTTGATAGAATCTTTTACTTTATCAACTAGTAGTATATCATA



## Annexes

GAGGGATTTTAGGGATAATTGCTTATGATCAACAAATGATTTTATACGTATTTTTGGGGATTTTTTTTGGAGAAAATATTCATGTTATATACC  
CAGATTTTAAATGGATGACTAGTATGATATTTCTTCGATTTGTGCTAATTAGGTTGTGCAATTTATGGTTTTTTCAGCTAAGCAATTTATCT  
TCGATGGAGGGTATTACTACTAAGCCTCACTTTGAGCAGCTGTTTCAAGGAGACCGGTAACCTTCAACAAGAACTTTCCAGTCACTATGTTTC  
TCATTTCTGCGGAGATTGATTTCTTTTCGATAACTCTATGTTTTTGGTTTTTAAATTTGCAATTTATTTTCATTTTCAACAGCTGCAAAAGTCAGCT  
GAGAAGTTAAATCCTGAACTGGTAAATGTTCCCTGCAACACTGTTTGTATTTGGCATTATTTTATAAATAGTGAATTTCAATTTATGGATTTTGA  
AATGCTGTGGGTGCTGATACTATGGCTGTGCTGTTTTTCAGACAAGGTCACCTAGCAAAGCTGCTAGCATGTTTTTCAGGGACACAAGAAAAG  
TCGCTACTTGGCGTAAACTGGTTATCCGCTCGAGAAGGTAAGAGTCAAGATTTCCAGATATTTTCGTGACATACATGCTTATAAGTTGCT  
TGATGGAACCTCAATAGTTAATTTTGATTGATTGGCTTGGTTGAGTTTTTCTGTTTCCAACGAGGCTGTCTGTGAATGCTCATTATGGAGCC  
TATATAACAAATTAATCTGAAGAAAAGGCTGGGGCTGTTTTGGGAAGTCAAAAGTGGCGGGCTACAGATTTAGAGAGCTAGGAATGACATGT  
AACTTAAATGCATTACTTAAGCTCGGGGTAATTTATCTGACATCAATTTGGACTTCGTGAATCAAGGAGGCTTGTATTGAAGTTCAATTT  
TTCTGGTGGACATTTGATGGAGGAGCTCGTATAACAACATTTGATAGAAATGCTGCTTATGACATACCCATGGAATAAAGGTTGCTCATATGT  
TATCTCAATTGACCATTCAAGTTGTTTAGCATCTGATAAATCACAATTCGTAATGCATGCTCCCTGTCTACTGCCTGAAATAGCCAGTTGTTT  
ATAGTGAATCCAAATTTTCTTTTCTTTTCTTTGTTGAGAAGAGCATAGTTATTTCTTGGTAGATGGTTTGTATGACTGTGGTGCATTCTACCA  
GTATTGTTATTTATCCATCCACCTCTCAGTCAAGGTTTTAGAAAACCTGGCCATTATATCAGATGCAGCATCAGTTATTTCCATTATAGAT  
TGGTGGTTGATAAAAAGCTTTTGAAGGAACAAGGAGGCTACCTTGCCTGGTCAAGTTTCTTCTATTAACCTGGAAGGAGGTTGCTCATATG  
TTAATGGCCCCCAGCTCCTCGTCTTTTAGATGTTTTCAAAAACAAATGCTAGATATTTCCAAGTATGTGCTGCTGAGCTGTTTGTGAAAATTGG  
TTGAATTTAAATTTCCATGAAAATAAATGCACTCTCTTTGAGCTAGAGTTCTTTATTTTTCTTTGGTGCAGTTCCTCTGTATATGCTAA  
AAGTACTCTGGAATCATAACAATGCTCTGTTCTCAAAAACACTCAGGTAACAGTTGAAAGCCAAGCCTATCAAAAATCATGTTTTCAAGTGC  
TCGACCGTGGCTGTGCTATAACACCCTCAAAATATGACAGCTTAGAGGCGATTGTTATGCAAGCATCACTTTCCCAACTTTTCAAGGAAA  
AGGGCAGCTACAACCCTCATCAAGTCCGCAACAATGAAGCGTGCAGCGGCTTCTGTGCCAGAGGCTTAAATATCTCCATTTGATATTTTATT  
TGTGTGGCTTTTACTACTTTTCAATGTTGTTTTCTTCCATGCCCTCCGATTTCCATTTGGAAGGAGTAGGCAAACTTGGTTCTGCCTGTGTTTT  
TTCTCTCTTTTCCAGCTATACTGGGCATGTTGTTGTTGTTGATGATAGATGGGTTTCAAAATCTCACAATGTAATAATTACTTGGCGGA  
GGAATTTCAATTTGAAATTTGGATTCTACTAGTACTTACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT  
CTCATAACAACAAGCGTCTACTCTTGTGACACGGATAAATTTATCATATGTTACGATAAATTTTCTTGTCTCCATCTGTCCCAATCTTTATCTG  
CAATCTTGAATGATGAAAACAGAACTTAAAGGAAAATTCACCTTGAATTTGGGATGGAAGATGCTAGTGGAGTAGCTAACAGTAAGTGC  
TGTTTAAAAACCTCAATCTGAGACGTGGCGGTCCCGATTAGTCTTTTCAAAAACCTCCGGTAAACATTACGCTACTCAGAAAACGACTCTC  
TTTTTCAGCCACCGTCTTGACACTTCTGATACCGTGTGAGCTGTGAAAAGATCAAGTATGAAAATCAAGAAAAGCGGATGGGGGAGCCAAA  
TTGCGGAATCAACACTCAGCTCTCCATGCCATCTTAATTCAGTATCTGCAAGCCATAAAATGTTTTGAACTCTAATTTCTGTGTGTTGTGTT  
TAATTCAGGTTGTCGAGACAATTTAAGCTCTCTCTCAAAAACCCCTACCCGCCCTGAAGGCGCACCAATCTCGATAAAATCTACTCTCG  
ATAAAAAACCCCTAGCCCTAACATTAATGCCATCACCAGAAACAACCTTTCAATCAAAGAGTTCACGTTAGCAGGTGCATTGCTCTCCTCCTC  
TCAATCCTCCACGAGGAATGCTCTCCTGATCCCAAGAACCTTCAATTTGAAGCTAACTTGGCTTTTCAACGAGCTATCGAAAAGCTTATGCT  
GAGATGATTTTGGTGACAGATATGAGAGGAGGATGGTGAGTTGTTGGGAGTGGATTGTGGTGGTTTAGTGTGTGAAGAGAAGAGATTGGTGAT  
TGAAATGATGCCTAAGGTTTTGCTTTTATGAAAGATGAATCAAAGAGAGTTCAATGATAAGAGTGTGATGGAGATGAGATTTCTGCCGCGT  
CGCTAGGGCTCTGTTGGATTCCGCGATGGTGGCTACTCATAGTTTAGATGTTTTGTTACTCAGGTTCTTGGTTAGGACTTTGGATTTAGGTT  
TTTATGTTGATTTTTTTTAACTGTTGATTTTTTATTCGTAGGTTGATTTGACCAAGTGGATGAACTTGTCTGTACAAGCAAGCGGGAAAA  
AATCAGATCAAAATGATGTTGAGATATCCAAATCCGACCACATGCAGATCAGTCCATGCGAACCATCGAGTA

>transcrit PtW LIM2b

TCGCTGCAAAAAGCGCCAAATCCCGCTCCCTTCGCTAGCCCTTTATCTTTAAACACCCCTTACCCCTTCAACTCTTCGCTTTCTTTT  
CATCTCTCGATTCACTACCTTTGATTCCAGATCGAAATATCCCATCTTTCCGATTTCAATCTCAATTCGGTTATCCAAAGGAGTTAACAGGA  
CAGTCACTCAGATCTAGAGGGTATAGTTTATATAGCTAGAAACGATGTCGTTTACTGGTACCCACAGAAATGCAAGGCTTGTGAGAAGACTGTG  
TATCCTATGGAACCTTCTATCAGCAGATGGGGTGTCTTACCATAAACTGTCTTCAAGTGTTTTCACTGCAAAAGGCACGTTAAAGCTAAGCAAT  
ATTTCTCGATGGAGGTTATTTGATTACTGTAAGCCTCACTTTGAGCAGCTGTTTCAAGGAGACCGGTAACCTTCAACAAGAAAAGTTCAGTCACTGC  
AAAGTCACTGAGAAGTTAAATCCTGAACTGACAAGGTCACCTAGCAAAAGCTGTGATGATGTTTTTCAGGGACACAAGAAAAGTGGCTACTTGC  
GGTAAACTGCTTATCCGCTCGAGAAGGTAACAGTTGAAAGCCAGCCTATCACAATCATGTTTCAAGTGTCTCGCACGGTGGCTGTGCTATAA  
CACCATCAAAATGATGACGACTTTGAGGGCGTATTTGATTTGCAAGCATCACTTTCCCAACTTTTCAAGGAAAAGGGCAGCTACAACCACCTCAT  
CAAGTCCGCAACAATGAAGCGTGCAGCGGCTTCTGTGCCAGAGGCTTAAATATCTCCATTTGATATTTTTTATTTGTTGGCTTTATCTACTTTT  
ATTGTTGTTTTCTTCCATGCCCTCCGATTTCCATTTGGAAGGAGTAGGCAAACTTGGTTCTGCCTTGTTTTTTTTCTCTCTTTTCCCGCAGCTA  
TACTGGGCATGTTTGGTTGTTGATTGATAGATGGGTTTCAAAATCTCACAATGTAATAATTACTTGGGAGGAATTTCAATGCAAAATTTG  
GATTCTACTAGTACTTACTTTT

>CDS PtW LIM2b

ATGTCGTTTACTGGTACCCAACAGAAATGCAAGGCTTGTGAGAAGACTGTGTATCCTATGGAACCTTCTATCAGCAGATGGGGTGGCTTACCATA  
AAACTTGGCTTCAAGTGTTTTCACTGCAAGGCACGTTAAAGCTAAGCAATTTCTTCGATGGAGGGTATTGTTACTGTAAGCCTCACTTTGA  
GCAGCTGTTCAAGGAGACCGGTAACCTTCAACAAGAACTTTCAGTCACTGCAAAAGTCAAGTGTGAGAAGTTAAATCCTGAACTGACAAGGTCACCT  
AGCAAGCTGTTAGCATGTTTTTCAAGGACACAAGAAAAGTGGCTTTCGCGGTAAGAACTGCTTATCCGCTCGAGAAGTAAAGCTTTGAAAGCC  
AAGCCTATCACAATCATGTTTCAAGTGTCTCGCACGGTGGCTGTGCTATAACACCATCAAAATTTATGCAGCACTTGAAGGCGTATTGATTGCAA  
GCATCACTTCTCCCAACTTTTCAAGGAAAAGGGCAGCTACAACCACCTCATCAAGTCCGCAACAATGAAGCGTGCAGCGGCTTCTGTGCCAGAG  
GCTTAA

>PtW LIM2b

MSFTGTQKCKACEKTVYPMELLSADGVAYHKTCFKCFHCKGLKLSNYSSMEGVLYCKPHFEQLFKETGNFNKNFQSPAKSAEKLNPETRSP  
SKAASMFSGTQEKCATCGKTAYPLEKVTVESQAYHKSCFKSHGGCAITPSNYAALEGVLYCKHHFSQLFKEKGSYNHLIKSATMKRAAASVPE  
A\*

**PtPLIM2a gène LG\_XIV:2231228-2232712**

>PtPLIM2a\_gw1.XIV.1147.1

TAGTTTGTGTTGGCATTGCATGTTACTGACATGAATATTTTTGTTTCCGATGATCTGTGGGTGGTTTTATTGCTGGCGCGCTGTGTATGTATA  
TATATATATATTTAAACCTCAACCTAATTAATCTGGTTCTTTGAAATGAAAATAAAAAATAAAATAAATAAATTTCTTACATCTGAAGTGAA  
GAGAACAATCCGAGACGCAACTCCGATAAGCTTGAATTTGTTATTGCTCATAAATGACACACCATCATTTAAACAGCTGGCTGGTTGGT  
GTGTACCTTTTTCAGACAAGTATAGTATATCTTAGAGGTAGCTGGGGTAACTGTTCGTAATTTGTTTAAAAATTTAAAAATTTATTTCA  
TCTTTAAAAATATATTTTATTTGTTTATGATTTGGAATAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA  
TAAAAAACAATAATCACACCTTAAATTTCCATCTTCACTTACTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA  
TTTTCAATTTTCAATGATTAATTTTTGTTTTCCCCCAATTTGTTTTCCCCCAATTTGTTTTCCCCCAATTTGTTTTCCCCCAATTTGTTTTTACT  
CCAATCCAACAGGTTTACATCCCAATTAGATAAATAACATTTTTATCTCATAGATGTTTGGGATCTCCAATCACGTGTCGCTTTACACAGT

# Annexes

CGATCCAAAGCAATTTTCTCCATTGCAAGAGAAGCCAAGAATTTGTGACGAACCTTGGTATAAAATAGTTTGTTTTGACAAGAGAGAGACTTCT  
TTGTTAAATTTCCATGTGTCTGAGGAGGAACGAAGAACAACCCAGGGTGTCTTTTCATTTGTTCCGGACACCTCTTAGATTCTCAGACAACCTTGA  
ACACAAAAATATTTCCAGGATCCAGGATCATTTTGAAGGATACAAGGAAAGGAAATAAAAAGAAAGAACAAATGGCATTTCACGGGAACCTGG  
ATAAGTGAAGGCTTGTGATAAGACTGTTTATTTTGTGATATGATGTCTCTTGAAGGGGTGCCTTATCATAAATCCTGCTTCAAATGCAGCCA  
CTGCAAAAGGAACCTTGTGATGTTTTCTCTCCCTTCTTTATGCTCTGCTCTCGTCTCTCTCTTTTATAGTTTTTTCTTTTATAAT  
CTTAGTGTATTTAAATGGGAATTTATATTTATCTAAAAAATTTGTGCAATAAGGGTTCCTTCAAGGCTAAAACATATATATTAATGTTCA  
TGTGTTACAAGCTAGATTGATCAACTGGGTTTTGGAAAAAGCAAGCCAAACATTAACATTTGGTGGAGTTTCCAATTCACGGATTAGTAATAGT  
TGCCTTTCATTGTTTTTTTTTTTGGCTTCGGAGTTGTGACAAATGATAATCCACAATGAAAATAATCTGCGCATGGCAGATGAGCAACTAC  
TCATCCATGGATGGAGTCTCTACTGCAAGACTCATTTTGAGCAACTTTTCAAGGAGGGTGGAGATTTTCAGCAAGAATTTTCAGAAAGGTGATA  
TCTTTTTTCTATGTGCTCTTATTCAATTTCTTATAGTACTGTTTCGATGAGATGCCATTGTTTGATTAATTTTTTCCCATATACATTAACA  
ATTTGACGACATTTTCTGCAGGAAGCCCGAGAGGACTCATGACTGTTACAAATATTATAAATCTCCATTTTAAAGTGTAGGGTGTGTTGG  
ATGTCCCATCAACTGGTTTATGGTGGCATTGGAAACAATCTGATGAATGATAATTTTCTGGGATATTTCAAAAATCCATTCAGTTTTCAGAT  
TTTATCTTTTCTCCATGACTTATAATAACACTATAAATTTCTGTGCATTTTACAGTCTAGGATCCCCAGCAAGCTATCTTCAGTGTCTGTGG  
AACCCAAAGCAAAATGCTCCGCTTGGCGCAAACAGTGTACCCATTGGAGAAGGTTTGTGTTTTCAGCTCTGAACTTAAATAAATCCATTAGGTCT  
AAAGACCTGTGTCTATTTATACCCCTACATTTATCAGATAAAATGACCAACATATTTCTTAGGTACATCTTATATTATAGATGATCTATCTG  
AGAAATGTTTTCAAGTCTCTGTTATTTAATAGGCATGCCTATGATAATCTGGAATAAGTACAGATGTGCTATAACCATTGATGACATAGTTAC  
CATGGAGGGAGAAATGTTACCACAAAGACATGCTTCAGGTGTGCTCATGGTGGATGCCACTCACACATTATCTTATGCTGCTCTTGATGGAGTC  
CTCTACTGCAAGGTCCACTTCGCTCAGCTCTTCATGAAAAGGGAACCTACAGCCATGCTCTGCATCAGCTGCACACAAGAGATCAAATTTCA  
CACCTCTGAACCTGGCCGGGTCAAACCAGAGGAAGGAGCTGCTGTGGAGGAGGAAAGCCAGAAAGAGCAATCTTAAACAAAAGCACATGTT  
TTCTTGCATTGCTATCTTTCAAGTGGCAGGACTGCTTGTGATTTTCATGTTGTTGAATATTATTGCCGGATTGTGTTCAAAGTTGTTTTTGG  
GATTTGACTATCTGTGTACAGATGCCTTGCTACAGGATATGGAATGCAATAGTTTTTACAAATTCCTCTCTGCAATTTAATACTCTGGC  
AACAGGATACAATAAAGAAACCCAAATCCTAGGAACCCAGTATCAACATTTATGATCTCAGCTCTTGTCTCAGATGGCTTGCCTTTCTATGGA  
TGCCAGATTTTCAACTCTCTACCATGCAAAATTCATTTCCAGCTCCAGCTCCCTGAAATGTAAGTGAATAAGCAAGAAATGTTGTTGAT  
TTCAAATATTATTTATTTATAATCAGAAAACTAAGAGGGTTAGTGGTTTCAATGTTGCCCTCTCAGATAAATTAATGTGGATTAGAATAGTAA  
TCCCTTCTTTTAGTCTTAAAGATTTTTTTTTTTTGTACTTTTAAATTTGTTTTGTTATTTTGGTTCTTAAATGTTGCATTGATTTGTGATT  
AGTCTTGAATAATTTGTTTTTCTATTCTTACTTATAGGATCTCATGTTGTAGAGAAATTTCTATCATTGAGTTGATTTCTAGTTTAGAGAAA  
TAAAATCAATTTTATTTAGTTTAAAAGAAATAGACCTCTTTTGAAGTCTTAAAGCTTAAAGAAATTTGATTTTTTAAATTTTAAATAAAA  
ATCAATGACTAATTTGTGTGATCAAGGGAAGAGCTCATGTCTTTTTTTGTGATTTGAAAATTTCTAACTCGGTTGATCTTTGATTTAAAAAA  
ATATAATAAAATTTATATTAATGTGAAACAATTAAG

>transcrit\_PtPLIM2a

CACAATTAGATAAATTAACATTTTATCTCATAGATGTTGGGATCTCCAATCACGTGTCCGCTTTACACAGGTCGATCCAAAGCAATTTTCTCC  
ATTGCAAGAGAAGCCAAGAATTTGTGACGAACCTGGTATAAAATAGTTTTGTTTTGACAAGAGAGAGACTTCTTTGTTTAAATTTCCATGTGTCT  
GAGGAGGGAACGAAGAACAACCCAGGGTGTCTTTTCATTTGTTCCGGACACCTCTTAGATTCTCAGACAACCTTGAACACAAAAATATTTCCCGAGT  
TCCAGGATCATTTTGAAGGATACAAGGAAAGGAAATAAAAAAGAAAGCAAAATGGCATTTCACGGGAACCTGGATAAGTCAAGGCTTGTGATA  
AGACTGTTTATTTTGTGATATGATGTCTCTTGAAGGGGTGCCTTATCATAAATCCTGCTTCAAATGCAGCCACTGCAAGGAACCTTTGTGAT  
GAGCAACTACTCATCCATGGATGGAGTCTCTACTGCAAGACTCATTTTGAGCAACTTTTCAAGGAGGGTGGAGATTTTCAGCAAGAATTTTCAG  
AAAGGGAAGCCCGGAGGACTCATGACCTGTCTAGGATCCCCAGCAAGCTATCTTCAGTGTCTGTGGAACCCAGACAAATGCTCCGCTTGGC  
GCAAAACAGTGTACCATTGGGAAGGTTACCATGGAGGAGAAATTTACCACAAGACATGCTTCAGGTGTCTCATGGTGGATGTCACATCAC  
ACATTCATCTTATGCTGCTCTTGTGAGTCTCTACTGCAAGGTCCACTTCGCTCAGCTCTTTCATGAAAAGGGAACCTACAGCCATGTCCTT  
GCATCAGTGCACACAAGAGATCAAATTCACACCTCTGAACTGGCCGGGTCAAACCAGAGGAAGGAGCTGCTGTGGAGGAGGAAAGCCAG  
AAGAGCAATCTTAAACAAGCACATGTTTTCTTGCATTGCTATCTTTCAAGTGGCAGGACTGCTTTGAGTTTTCATGTTGTTGAATATT  
ATTGCCGATTTGTTTCAAAGTTTGTGATTGTGATCTGTGTTACAGATGCTTCAGGATATGATGATGCTTACAGGATATTGAATGCAATGTTTTACAA  
ATTCTCTCTTGAATTTAATAACTCTTGGCAACAGGATACAATAAAAAACCCAAATCCTAGGAACCCAGTATCAACATTTATTGATCTCAGCT  
CTTTGCTCAGATGGCTTGGCTTTCT

>CDS\_PtPLIM2a

ATGGCATTTCACGGGAACCTGGATAAGTGAAGGCTTGTGATAAGACTGTTTATTTTGTGATATGATGTCTCTTGAAGGGGTGCCTTATCATA  
AATCCTGCTTCAAATGCAGCCACTGCAAGGAACCTTGTGATGAGCAACTACTCATCCATGGATGGAGTCCCTACTGCAAGACTCATTTTGA  
GCAACTTTTCAAGGAGGGTGGAGATTTTCAGCAAGAATTTTCAGAAAGGGAAGCCCGAGAGGACTCATGACCTGTCTAGGATCCCCAGCAAGTCA  
TCTTCAGTGTCTGTGGAACCCAGACAAATGCTCCGCTTGGCCAGCAAAACAGGTATCCCATTTGGAGAAGGTTACCATTGGAGGAGAAATGTTACC  
ACAAGACATGTTCCAGGTGTGTTTAAAGTGTGCTCCACTGCACACATCTTATGCTGCTCTTGTGAGGAGTCCCTACTGCAAGGAGTCCACT  
CGCTCAGCTCTTTCATGAAAAGGGAACCTACAGCCATGCTCTGCATCAGCTGCACACAAGAGATCAAATTCACACCTCTGAACTGGCCGG  
TCAAACCAGAGGAAGGAGCTGTGTGGAGGAGGAAAAGCCAGAAAGAGCAATCTTAA

>PtPLIM2a

MAFTGTLDKKACDKTVYFVDMMSLEGVPHYKSCFKCSHCKGTLVMSNYSMSDGVLYCKTHFEQLFKEGGDFSKNFQKGPERTHDLSRIPSKL  
SSVFCGTQDKCSACGKTVYPLEKVTMEGECYHKTCFRCAHGGCPLTHSSYAALDGVLYCKVHFQALFMKGTYSHVLSAAHKRNSNTPPELAG  
SNPEEGAAVEEEKPEEQS\*

**PtPLIM2b gène LG\_II:11974041-11975746**

>PtPLIM2b\_grail3.0039003401

ATTCTATCTCAAGGCTAATATTTGAAATAATTTTCAGAGGACTTAATTCGTGTCAACAAACAATTCGAACAAATTCGGGAATAGAAAATATGTTA  
TTGGAAGCATGPTAATTTACAGGGGAATGTTTCAGCAAGACTAGACTAATTTCCACCTAATAGAAATTTTATTGTTTTTTTTTCTAATAAAGACA  
AATTAGCTCATTATACTATTTAATCTACACTACAATGAGTCTGATGCGCAGTGAATTTTTTTTTTTTTTCAATACTGAACGGGATGAGTTAA  
AACCTGTTTTATCTTCAAGATTTGAAAGATAAAGAAAAGCATGTCGCCCTGTATATAAATTTAAAGATCAAATTTAGTTTCTAGATTGCAAGTCAA  
AGACAGCATGT  
TCTAATCTGAATGTCAGTTTGAAGTTAGAAAAAGAAATCCAATCTTGAAGTGGCGAGAAACAATCCAGAGCAGAGCAGAGCAGACCACC  
AATCCCAACCTGTGATAATGTTAGAATCCAGAGGTTTTGACTTTTGTACTGCTGTAATGCCAATTTACTAGGAACATTTGATTCACATAAGGG  
CCACAAAAGGGGTTGGAACAGCATTTAAACAGTCTGGTTTGGTGTGATACATAGACAAATTTGATAGCATATCTAATCTTCAACTTAAAAAAA  
AAATCAAGGGAATTTACAAAAACAGTTGGTTTTTCAATTTTCTTGTATAAAGATTTTCCATTTTGCCTCCCTTACAATTTTCTCAAGAAAA  
GTCAGAGCAGACAGACTTTTTTACTCCAATCCAACAGGTTTACATCCATAATAGCAGTCAATTTATATCTCAGCATGCTTGGGATCTCCAATC  
ACGGCTCATCTTATACAGCTCAACACAAAGCAATTTACACATTTACACGAGAAGGCAAAATTTGTGATAAAGTATTGTTTTGTTTTGACAAA  
ACAAAGACTCTTTGTTAATTTCCATTTGCTGTGAGGAGGAGCAACAATCCAGGGTGCCTTTTATTGTTAGGACATTTCTTGGTTTCTC  
AGACAACCTGAACAAAAAATACTTTCTGACTTCCAGGATCATTTTGAAGAGACGAGGAAAGAGGAGCTAAGGAAAGAAAGCAATGGCATT

## Annexes

```
CCTGGAACCTGGATAAGTGAAGGCTTGTGACAAGACTGTTTATGTTGTGATATGATGCTCTTGAAGGAGTGCCTTATCATAAATCCTGC
TTCAAATGCTCCCATGGCAAAGGAACCTTGTGGTATGTTTCTTCTGATCTTCCCTTCCATAGCCTTCTTGTATAATCTTAGTGTATTATA
GAATGGGAATTACTTAAACAACATGCAATGGGTCTAATTTGAAGGCCAATAGATAGCTAAAAATTTAGTAATCTTCAACACAGCAAGCTAAC
CTAGGATCAATGGGTTTGGAAAACAGCAGGCCAGCATTCAACAATGGTGTCAATCGCCGATAGTGATTCAATGCTTTTTCATTGTTTGGCTT
AATCTGGATCTGTGACAATGATATCCACAATAAAAAATAAAATGTTCAATGATTTGAAAATGATCTGTGCTTGGCAGATGAGCAACTA
CTCATCCATGGATGGAGTCTCTACTGCAAGACTCATTTTGAGCAACTTTCAAGGAGGGTGGCGATTTGAGCAAGAATTTGAGAAAGTGAT
ATCTTTTTCTCAGTGTCTTTTATGCAATCTCTTACAGTATTTCAACAGAGTTGCCATTGTTAATTTCTTCCATACACATTTAAACAATTTA
TGACATTTTCTTGCAGGCAAGCTGAGAGGACTCATGAGCTGGTATGATATTCGTACATATATATCGTATCATAATTTAAAACTTGGGGCGT
GTTTGTATGTCCCGTAAACTGATTTATGGTGGCCTTTGGAAACCACCTGATGAATGATATATCTGGGATATTTCAAATCCATCCAGCTT
TCAGATCTGTATCTTTTCTTCAAGATTTATAATGCTACTATATTTCTCTGTATTTTACAGATTAGGACCCCAAGCAAACTCTTCCGTG
TTCTGTGGAACCAAGACAAATGCTCCACTTGGCGCAAAACAGGTGATCCCAATTTGGAGAAGGTTTTGTTTTTAACTTAACTTGAATAAAATACAC
AGGTCTTAAAGATCTATGTCTTATACCTCTACATTTATCATACGAATGACAACTATTCTGAGATACATCTTGTCTATCATAGACATGGACCTA
TCTAAGAGTTTTGATAAGTGTATTTTAAACAGTCTTGCCTTATGATAAATTTGGAATAAGCATGTTTGAACCCGTTTCATGACATAGTTACCAT
GGAGGGAGAATGTTACCACAAGACGTGCTTCAAGGTGTGCTCATGGTGGGTGCCACTCACACACTCATCTTATGCTGCTTGGAGGAGTCTCT
TGCCTCAAGGTCCTTGTCTCAGCTCTTCAATGGAAAAGGGAACCTTCCCTCGCAGGAGCTACACATAAGACATCAACTCCACGC
CACCTCTGAACTGGCCGAGTCAAAACAGAGGAAGCAGCTGCTGCTGAGGATAAGTCAGAAGAACAATCTTAAACGAAGCAGCACATCATGGT
TCTTGCAGTCTATCTGTTCAAGCTGTCAGGACTGTTTTGAATTTTCAATGTTGTTGTAATATTATGCGGATGTTGATCAAAGTCTATTTTGTG
GATTTGATTAACTTTGTACAAATATGCTTGGTACAAAGATATTGAACTGCACTAGTTTTTCCAAATCTTCTCTTGAATTTAGTACCCTT
GGCAAACTGAGAAAATATGAAAAGATCTCAAGTAGTAATATCAACATGATTGAATGATTGTTCTTCTCGGCTATTTGCTTAAATGGCTTCCG
CATTTCTGTGGATGCCAGAGTTTGAACCTCTTACCATGCAAACTTATATTTCCAGGAACCTTGAAGATTTTAAACCTTTTGAACATTTGAAGAAGA
ACAAAATTAGGTGAGTTTGGCTTAAATGTTGTATACAAGTTTCTTGAATTAAGAGATAAATAGAAATTTCAAATTTCTTCTTATTTATTT
CATCAAGAAGCAACCAAGCAGGGGATTTAAATCTGCTACAAATCCCAGAAAACCAATACAATAGACAATCTGTATTGAAGACCAATCCATAAAA
CAAAACATGACCAATTTGAAATTTGAAATTTCCACCTGCTACAAAGATCAGCGAATTTAGCATGCAACATCAATCAATTTGA
ACAATATGAGCAACTACGGCAGTGGCTTGGGAGCTTTCTTCAATTAACATCTTGGCATTTCCCTCTTCCAAACAGAATAATGACCCACACCA
ACGAAGTCTGAGGATAACTGAAGAAAAGCACTTCTTCCATAAATTTCTTGTAGATAAACAACACATTTGCCAGTCTCGTACAGTCTCTTGAAT
TCAATAAAACCGCAAAATAGAATCCAAATATCTTTAGCAAACGAGCTAAGTAAAAAACAGATGATCCCTATCTTGTAGTGCATAGAACACAT
AGAGGATCGATATTTCTCCCACTACTGCTGGTCCGACCT
```

>transcrit\_PtPLIM2b

```
AATTTGTGATATAAGTTAGTTTGTGTTTGAACAAAACAAAGACTTCTTTGTTTAAATTTCCATGTGCTGAGGAGGGAGCGAACAAATCCCAGG
GTGCTTTTATTGTTAGGACATTTCTGGTTTCTCAGACAACCTTGAACAAAAAATACTTTTCCGACTTCCAGGATCATTTTGGAAAGAGACGAG
GAAAGAGGAGCTAAGGAAAGAAAGCAATGGCATTCACTGGAACCCCTGGATAAGTGCAAGGCTTGTGACAAGACTGTTTATGTTGTGATATGAT
GTCTCTTGAAGGAGTGCCTTATCATAAATCCTGCTTCAATGCTCCCATTTGCAAAAGAACTCTTGTGATGAGCAACTACTCATCCATGGATGGA
GTCTCTACTGCAAGACTCATTTTGAGCAACTTTCAAGGAGGGTGGCGATTTGCAAGAATAATTTGAGAAAGGCAAGCTGAGAGGACTCATG
AGCTGATTAGGACCCCAAGCAAACTCTTCCGTGTTCTGTGGAACCAAGCAAAATGCTCCACTTGGCGCAAAACAGTGTACCCATTTGAGAA
GGTTACCATGGAGGGAGAATGTTACCACAAGACGTGCTTCAAGGTGCTCATGGTGGGTGCCACTCACACACTCATCTTATGCTGCTCTTGGAC
GGAGTCTCTACTGCAAGGTCCTTTGCTCAGCTCTTATGAAAAGGGAACCTATAGCCATGCTTCCGAGGAGCTACACATAAGAGATCAA
CCTCCACGCCACCTCTGAACTGGCCGAGTCAAAACAGAGGAAGCAGCTGCTGCTGAGGATAAGTCAGAAGAACAATCTTAA
```

>CDS\_PtPLIM2b

```
ATGGCATTCACTGGAACCCCTGGATAAGTGAAGGCTTGTGACAAGACTGTTTATGTTGTGATATGATGCTCTTGAAGGAGTGCCTTATCATATA
AATCCTGCTCAAATGCTCCCATTTGCAAAGGAACCTTGTGATGAGCAACTACTCATCCATGGATGGAGTCCCTACTGCAAGACTCAATTTGA
GCAACTTTTCAAGGAGGTTGGCATTTTCAAGAAATTTTCAAGAAAGGCAAGCTGAGAGGACTCATGAGCTGATTAGGACCCCAAGCAATTTGAG
TCTTCCGTGTTCTGTGGAACCCCAAGCAAAATGCTCCACTTGGCGCAAAACAGTGTACCCATTTGGAGAAGGTTACCATGGAGGGAGAATGTTACC
ACAAGACGTGCTTCAAGGTGCTCATGGTGGGTGCCACTCACACACTCATCTTATGCTGCTCTTGGAGGAGTCCCTACTGCAAGGTCCTT
TGCTCAGCTCTTATGAAAAGGGAACCTATAGCCATGCTTCCGAGGAGCTACACATAAGAGATCAACCTCCAGCCACCTCTTGAAGTGGCC
GAGTCAAAACAGAGGAAGCAGCTGCTGCTGAGGATAAGTCAGAAGAACAATCTTAA
```

>PtPLIM2b

```
MAFTGTLKDKKACDKTVYVVDMSLEGVPYHKSCFKSHCKGLVMSNYSSMDGVLYCKTHFEQLFKEGGDFSKNFQKGPRTHELIRTPSKL
SSVFCGTQDKSCSTCGKTVYPLEKVTMEGEBYHKTCFRCAHGGCPLTHSSYAALDGVLYCKVHFAQLFMEKGTYSHVLAGATHKRSTSTPPPELA
ESNTEEAANAEDKSEEQS*
```

### PtδLIM2a gène LG\_I:7976496-7977725

>PtδLIM2a\_gw1.I.9051.1

```
GGAGAGGCAGAGTTTGTATGTGAATGAGATGTCAAAGACTGAAACCTTTTTCGAATTTCAATTTTTCAGATAAACTTTTTCATGATCAATATTCGC
TACATTTGTGAGCTGATAATAGATAATTCATTTCTACTTGTAGCATAATCAATTAATGTGAGGCTAATATCATCATGTCAACAGGAGATACGA
GGCCTAATATTTAGCCCAATAATTTTCTTTATTTCCCTTTTCCAAAGTAAATAAGAAATAAGTATGAAAAAATTTCCAAAGAACTAATCAATGATC
CCTAATGCCTTAAATCTTTATATGATGCCACGGCTGCAGAAGAAAATCTGTGCAATTAACAATTTATCCGTTTATTTTGAATACTTTCAGCACT
CAATATAATCCTAGTACAGACCAGCTGCTGTGCACATTTGCTGAGCATTTTATATATATAGAAATTAAGCAGAGACTCTTGAATTTGTTGGCCTA
CAAAATTTGGAATCAAATTTGATGCTAACATAGTATTTTGTGTCATGTACACAGATTTTTCAGAGATAGGCTGCTAGATATAGTAAAAAATCT
ATTTGGAGACAGCTGAGTTGCTTACCACATAAACAAGAGGATGCGATTTGATTTGTTTACCTTCTGCTGACTATCTTTGAAAAGACTTGGTCAG
GAAATGTTTCAGATCTATCTGTAACAGCCACTATTCTTGAACGGTACCCTGCAAGTCCAAGAGATATAGACCCTGGATCACATTAGCTTAA
ATAAGCGGTTCAAGTCTTTTCTCCATCCATTGAACATTCATTTTATGACCTTTAATCTGCTAGTTAATAACAGGAAAGGTTAAAGGTTAAGA
ACAGATTTTCAAGTCTGTTTATCTTAACTCTGAGATTAAGAGCCGAACCTTCAAGATTTGTTCTTAAACTTTCCCATTAAGCTTTTCTTA
AGCAACCCCTTCTGAAATCAGATTTCTAATCTACTGCTCATTTTAGATTCATCAGGAAAACATAATGGGATTTCACTGGGACTTAGAGAAATGGAAGG
CCTGCGACAAGACTGTTTATTTTCATCGAATTTGGTGTCTGCTGATGGAGTTCCTTATCATAGAAGTGTCTCAAATGCAGCCACTGCAATGGACT
TCTTGTGTTACGAGTCTTATTCATTTATCCTCTTTTCTTTTTCATGCCTTTTCTTTTGTGGTTATTTGCTAAGTAAAGAAATTAATGAAAATACGG
TCCTTTACCAGGATTAATGAAACTGACCACCATGAAATTTGCTTATAGATGAGTAGCTATTTCTCGATTGATGGAGTTTATACTGCAAAAC
TCATTAACGACCACTCTTCAAGGAGACCGGTAATTTTACCAAGAAATTTCAACCTGTGAGATCCGCTTCTCCGTTCTTCTTCTTCTTAAAGTTT
CTTTTACATATTTTGGAAATTAGAGTTCATGGAGAAGGATGTAGTCCCTCCTATTTTACCTTTTTTTTTTTTTTGGCTAACATGTGTTTCTGCA
ATTTCTGACTGTTTAGATGAAGAGAAGAAAACCTGCCTGTACCAAGATCTTTACAATCGGGTCTTAAATTTTTTTCATTACCTAATTTGTTTCTA
ATGGAAGTTTCTAGTTTGTGTTGCTGGAGTTGATGCATGTTTTCACAATCATGCAGACAAGGCCCAAGCAAACTTCTCTCATGTTTCACTGTC
CACCAAGACAATGCTTCTTCTGTAATAAAACCTGCATCCCATTTAGAGAAGTTTCTAATCCATCTCTCTCTCTCTCTCTCACACACACAC
ACACACAATTTCTATTTCCCATTCATTAACATTCATTTCTCTGTCATATATATATTTGCGCAACTCGAATGGAATGCAAAAGACAGAT
```

## Annexes

AACCTCTCATTGATATCTATATTGTACCTTAATCTGGCACCAATATAAATCTCATGCACCAAGATCAACTCCGGAATAGATTGAGATCAATATTA  
TATGATGGAGTGAGGATAAAGAAGCTTATTGCCGTAACCAAAATTTTGACAGGTGACTGTGGAGGGAGAATCTATCACAAAGTCATGCTTCAGG  
TGTCTCATGGGGTGTCTGTATCACACCTTCTTCTTATGCGACTCTTGATGGTATCCTATACCTGCAAGGCCACTTTGGCTCAATTTGTTAAGC  
AGAAGGGTAGCTACAGTTATCTCACCAAAACCTGCCACAATGAAGAAAACCGCAGTAAATTTGCCAGAAGAAAATCCGAAGCAGCAGAAAACAA  
TGAACAGTACCAGAAGCAAACCCAGATGCTAACTCTGATTTAGCCATTGACAAGAGACATCATAAATTCATGAATCTCGACAGCTAGTTCTC  
AATGCTGTGTTGTTTCCTTTCACATGCCTCTCTAGGGACATTGAGGTTACAAGAGTGGTTGCTTTCCTTGTCTTACTCAGGCATGTGTCAACA  
GAGAATGGGAACCTGGAATTTTGTTCGCCTTGAATTTGGAACATAAATAGAGACTTCTGTAATATTTCTATTTGCGAGGTGCAGTTGATGAA  
CGAAAATAAAATTTCACTTCTTTCTCTTTTGTCTTTAAATTTCTTTCATCAATAAAAAAATGGTCTTGAAGATGAACAAAACCATATTAACA  
CGGGAGAATCTAGACATGTTAAAGAACGGTCGACGACATGAGTTTCTTTGGCTCTTACACTGCTACAAGACTGACTTTAATTTTTCCATA  
TTCATTTGTCATTGGAATTACAAAGTTGTAAGTGCATTTGCAACCTTTTTTTTCAAGTAATGATACATGAAATTTGGATGAATTTGAGCC  
CCCTCTGGAAACCGCAGTCCATAAGTCATGCTCCAACCTTTTGTATAAATCTTGCAAGTGCATAAAATTTCTCAGATTAAGCTTGATATTTTCC  
TTAGTGTACACTCAATTGATTGTAAGCATGTGATTTCAAGCACACCAAAATCTCCACCTAAAACCTCTTCTTATATATCTGCCACTTGT  
CATCATGTCCCTGGCCTTTTCCAGTTAAAATTTGCAACCATTTGATCCAATCTAGCCCTTATTGAAATCACCTTCTTAATGTTGCTACTGGAGTAG  
TTGCTTGAAGATTAACCTCTACATTTGGGTGAAGGATAACCATGCAATGATTGAAAATGGCTCATTTTAAATGGATAAATGATAGCTAGAATTATA  
CCATCATTTGAAACTAAAAGACAACCATTTGATCATCTGGATTATGCATGATCATACACCTTAAAATAGGTTTTAAGACATTGATTAACCCCT  
TTCTGTTTATCTGTCATTTAAGGATAGAAAGATGAGTT

>transcrit\_CDS\_PtδLIM2a

ATGGGATTCACCTGGACCTTAGAGAAATGCAAGGCCTGCGACAAGACTGTTATTTTCATCGAATGGTGTCTGCTGATGGAGTTCCTTATCATA  
AGAAGTGTCTCAAATGCAGCCACTGCAATGGACTTCTTGTGATGAGTAGCTATTCCTCGATTGATGGAGTTTTTACTGCAAACCTCATTACGA  
CCAGCTCTTCAAGGAGACCGGTAATTTTACCAAGAAATTTCCAACCTTATGAAGAGAAGAAAACCTGCCTGACAAAAGGCCCAAGCAAACCTTCC  
TCCATGTTTCAGTGGCACCAAGACAAATGTGCTTTCTGTAAAAAATGCATACCCATTAGAGAAGGTGACTGTGGAGGGAGAATTTATCACA  
AGTCATGGTTCAGTGTCTCATGCGGTTGCTATCACACCTTCTTCTTATGCAAGTCTTGTGATGGTATCCTATAAGGCCACTTTGCG  
TCAATTTGTTAAGCAGAAGGGTAGCTACAGTTATCTCACCAAAACCTGCCACAATGAAGAAAACCGCAGTAAATTTGCCAGAAGAAAATCCGAA  
GCAGCAGAAAACATGAAACAGTACCAGAAGCAACCGCAGATGCTAACCTGATTTAGCCATTGACAAGACATCATAATTCATGA

>PtδLIM2a

MGFTGTTLEKCKADKTVYFIELVSDGVPYHKKCFKCSHCNGLVMSSYSSIDGVLVYCKPHYDQLFKETGNFTKFKQPYEEKKNCLTKAPSKLS  
SMFSGTQDKCAFCKKTAYPLEKVTEVEGFYHKSCFRCSHGCCIPTSSYAALDGLIYCKAHPAQLFKQKGSYSYLTKTATMKKNVNLPEKSE  
AAENNETVPEANADANSDLAIDKRHHNS\*

### PtδLIM2b gène LG\_III:11773343-11774908

>PtδLIM2b\_gw1.III.1568.1

GTTAGGACATGGGTGAGCATTCTGTTAATTCAAAAACAAGTTCATTTCAACATGAAGCCGTTTGGCTTGGTATGTGGTGTGTTAGGGGCTCG  
AGATTTGCTTTGGAATTTTACGGAATGGAAGTGTCTGCTGCTAATTTGGCAGAACGAAGATGACCTGTAGATGTTTGTAGGCAAGTTTATGC  
GATCTCCTCGTATCTTTTTTATAATGGCAACTGAGCAGTTAAAGAAATTTATGACAGTAATTTTTCAGGGACATTAATGATTTTTCTTTATTTTT  
TTTTAACACCATGTAAGTATTTTGAACCTATAAAGTATTTGGAATAATTTTTCACAAACTTTTCAAGAAATATGAAAGTTTTTTTCTCTTATT  
TCGAGGTATAAGTAGCGTCAGTGTATTGCAACCTATCCATTTATGAATATACATCATGTCCTGGAACAATGTTTGCATCTTTTATGTAGGATAT  
AAGCAAAGACTCCTGAGTGGTTGGCCGATCAATTTTGGAAATCAATGGATGCTAATATAGTGTCTTATTGTCATGCCACACACGAACTCTGAA  
GTTACAGATTTGGTTGCTAGATATAGTAAAAATCTATATGGAGATAACTGATTTGTATACCACATAATAGGAGGATTCAGTTGTAATTTGTGTC  
CTCTGTGAGGACCTTTTAAAGACTTGGTCAGGAAATGCTTCATGCTTATCTTTTATACCCCACTATTTTGAAGTCTTATGATTTGATTTTTC  
CAAGAGACATGAACCTTAGTACAATAGCTTAAATAAGCGATTTCATGTCCTTTTCTCCATCCATTGAAGATTTCATGTTGATCTGGGTCAG  
ATATTCTGGGAATTCATCTATCTTAACTTCTGAGATAGAGTGAACGGAACCTTCAGAAATTTGTTGTTGAACCTACTTTTAAAGCTTTTCCA  
AGCATCCCTTCTGATATCAGATTTTATACTGCTGTTTTAGATTCACAGGACCAACATAATGGGTTTTCACCTGGGACCTTAGAGAAATGCAAGG  
CCTGTGATAAAAGCTTTTATTTTCATTGATTTGGTGTGCTGCTGCTTATCATAAAGAGTGTCTTCAATGAGGACCTGCAAGCTGCAATGGACT  
TCTTGTGTCGAGTTTCTATTTCATTTCCCCCCCCCCCCCCCCCGCGATGCCTTCTCTTTTGTACCTGTTGTCCATGTAAGAATTAATGAAT  
ACGATCTTTTACCAGGATTAATGAATTTGACCCACCATGAAATTTTCTTATAGATGAGCAGCTATTCCTCAATCGATGGTGTGTTTATATTGCA  
GACCTCATACGACAGCTCTTCAAGGAGACTGGCAATTTTAGCAAGAACTCCAATCTGTGAGATCCCCCTCTTTGTTTCTCTATGTTCAA  
TCTTTTACATACTTTTGAATTTGGTTTTCCATGGAGAAGTGTAGTCTCCCTAATTTTACCCTTTTGGCTTACCAATGTTGTTCTGCTG  
ATTTCTGGCTGTTAGCGGGAGAGAAGAAAATGGCTGTACTAAAATTCGCTTACAACCTGATTTTTCATTACCTTTTTTCTTTTCTGATGG  
AAATCTCTAGGTTGTTTGTGTAAGTTTGTATGATGTTTTCCCAATTACGCAGACAAAGGCCCAAGCAAACCTCTCTTCCATGTTCTCTGGCACC  
CAGGACAAATGTCATCTCTGTAATAAACTGTATACCCACTAGAGAAGGTTTCTTATCCAATTCCTTTCTCTCTCAGAACTTATTTCC  
CAATTAATTAACATTCATCCATCCATTAATGATATATATGATATATTTGCAAACTCAAATGGAAATGCAAAAGTCAGAATAACTCTCATGATATAT  
TTGTTGACCATACCTGAAATCAATATAAATTTGACGCAAGTAAGAAATATTTTCAATTTAATAAATTTAATTTAATTTAATTTAATTTAAGTT  
AAGATAAAAAAATTTATATAATATAGAGTCTTGTCAATCAAGTATTACATGTTTGAATTTCAATCTTCTCTTCTAATATAAAAACTAAAC  
ATGAGATAGTGTGAATATATACAAATTTCAAAATTTAAATAAATTTGTGTTTGAATAAATATGTTAAATAACAATAAATAAATTTATCTTGACACCT  
CATCTAACAACTTAAGTTTTAAGTAAAAATTTTTTGTATATAGTATTAATAATTTTATGATTAATTTGACTATAAATTTAAATTTATCATATAT  
TCATTTATTTTATAAGTTTCAGATGGTTCAATACATTTGAGATAGAAAATAATATGATCAAGTGAAGTAAAAAGGCTTATCGCTGTAACCAAAA  
TATTCTTGACAGTGACTGTGGAGGAGAATTTCTTACAAGTCATGCTTCAGGTGCTCTCATGGAGTTGCTTTATCACACCATCATCTCAGC  
CTGCTCTTGACGGCATACTATACCTGCAAGGCCACTTCTCGCAATGTTTAAAGCAGAAGGGAAGCTACAGTTATCTCACCAAAACCTTCCAGAT  
GAAGAAAACCGCAGTAAATTCGCCAGAAGAAAATCTGAAGCAGAACAAACCATTTAACAGTACCAGAAGCTAGCTCTGACTTAGCCATTGCA  
CATGAGAAATGTCAGAACTAGTTATCAATGCTGATTGTTCTTTCACATGCCCTCAAGGGGACATTGAGGTTACAGAAGTGGTTGCTCTCCT  
TGTTCTTGAAGGCATGTGCAATAGAGAATGGAACTGTTAAACAAAATAAATGGTAAAAAGAAAGTCTTTTTATTTATTTTTTCTTTGGCC  
ATTTCTCTCTAATAGTGGTTTCAACAAACAAAAAAGGATCTCACTTGTATCAATATAAATCTTCTCCACATGAAGGATTTTAACTGTTA  
GACGACATCTACCTGTAATTTCTATAAGATGTGAATTTTACATTAATAAACCATAAATCAGTTCGTGAGGCGTTAATGATAATTAATC  
ATATCCAAAGAAGCATAGTTTATGTTGGACATATAAATTTATGTTGATATTAACATAAACAATTAACAATTAACAATTAACAATTAACAATTTTACT  
TGAATAACTAGCCATTGCCTTGTGTTGCACCGGGTGGTTTTTTTTCTTTTTTATAGTTTAAAAAATATATCTTTTTATGTTTTAAATTT  
GATTTATTTAATTTGTTTCTTAATCTTTTATAAAAAATATTTTTTTTTAATTTTACCCTTCAATCAAGAATTTATTTTATGATTTTATGTTA  
ACGTTGATTTCTTTTTATTTTTTTTTTTTTTGTCTTTTACTAAATTTGATTTTTTTTTAATTTCTATCTTCCATCAATATAAAAAATTTATTA  
TGATTTTTAATTTGATCTTTATTTTTTAAATTTGTTTTAATCTTTTATATAAATTTGAAATTTTTTTTTTCTTATTTTATTTTATTTTGT  
ATGATTTGAGAATGAACTCATGATTTTTCTAGATAGGGTGTCTCAAGTCTAATAAATAGGTCAGGTTTGAAGGTTAATACGAA

>transcrit\_CDS\_PtδLIM2b

ATGGGTTTTCACCTGGACCTTAGAGAAATGCAAGGCCTGTGATAAACTGTTATTTTCATTGAATGGTGTGCTGAGTGGAGTTCCTTATCATA  
AGAAGTGTCTCAAATGCAGCCACTGCAATGGACTTCTTGTGATGAGCAGCTATTCCTCAATCGATGGTGTGTTTATATTGACAGCTCATTACGA

## Annexes

CCAGCTCTTCAAGGAGACTGGCAATTTTAGCAAGAACTCCAATCCTCGGGAGAGAAGAAAAATGGCCTGACAAAGGCCCAAGCAAACCTCTCTCCATGTTCTCTGGCACCAGGACAAATGTGCATCCTGTAAAAAACTGTATACCCACTAGAGAAGGTGACTGTGGAGGGAGAATCTTTTCACAAGTCATGCTTCAAGTGCCTCATGGAGGTGCTTTATCACACCATCATCTACGCTGCTCTTGACGGCATACTACTGCAAGGCCCACTTCTCGCAATTGTTTAAAGCAGAAGGGAAGCTACAGTTATCTCACAAAACCTCCACGATGAAGAAAAACGAGTAAATTCGCCAGAAGAAAAATCTGAA GCAGAACAAAACCATTTAACAGTACCAGAAGCTAGCTCTGACTTAGCCATTGCACATGAGAATGTGCAGAAGCTAG

>PtδLIM2b

MGFTGTLEKCKADKTVYFIELVSADGVPYHKKCFKCSHCNGLLVMSYSSIDGVLVYCRPHYDQLFKETGNFSKLLQSSGEKKNGLTKAPSKLS  
SMFSGTQDKCASCKKTVYPLEKVTVEGEFFHKSCFRC SHGGCFITPSSYAALDGI LYCKAHFSQLFKQKGSYSYLTKTSTMKKNVANSPEEKSE  
AEQNHLTVPEASSDLAIAHENVQN\*

### Annexe 2. Informations sur les gènes codant pour des protéines à domaine LIM exprimés chez les plantes

Ce tableau représente la liste des gènes identifiés chez les plantes, nombre d'ESTs correspondant, les origines tissulaires des ESTs, et les ADNc, EST ou contig sélectionnés pour les analyses phylogénétiques réalisées à partir des séquences protéiques déduites.

Nom du gène	Espèce	Nbre d'EST	Type de tissu	ADNc, CDS, EST ou contig sélectionné
<i>AaW LIM2</i>	<i>Acorus americanus</i>	4	Inflorescences avec bourgeons de 10 mm	contig
<i>AfpaLIM1</i>	<i>Aquilegia formosa x A. pubescens</i>	3	Mélange de méristèmes apicaux de tiges et de fleurs; bourgeons floraux; feuilles et racines	DR924381
<i>AfpβLIM1</i>	<i>Aquilegia formosa x A. pubescens</i>	8	Mélange de méristèmes apicaux de tiges et de fleurs; bourgeons floraux; feuilles et racines	DR922857
<i>AfpW LIM2</i>	<i>Aquilegia formosa x A. pubescens</i>	14	Mélange de méristèmes apicaux de tiges et de fleurs; bourgeons floraux; feuilles et racines	DT742651
<i>AfpδLIM2</i>	<i>Aquilegia formosa x A. pubescens</i>	5	Mélange de méristèmes apicaux de tiges et de fleurs; bourgeons floraux; feuilles et racines	DR922206
<i>AmW LIM1</i>	<i>Antirrhinum majus</i>	8	Plantes entières	AJ792675
<i>AmβLIM1</i>	<i>Antirrhinum majus</i>	5	Plantes entières	AJ805529
<i>AmW LIM2</i>	<i>Antirrhinum majus</i>	7	Plantes entières	AJ802132
<i>BdW LIM2</i>	<i>Brachypodium distachyon</i>	2	Racines	DV487949
<i>BnW LIM1</i>	<i>Brassica napus</i>	8	Jeunes fleurs et bourgeons floraux	<b>DQ242645</b>
<i>BnW LIM2</i>	<i>Brassica napus</i>	3	Jeunes fleurs et bourgeons floraux; graines	CX192914
<i>BnPLIM2</i>	<i>Brassica napus</i>	1	Jeunes fleurs et bourgeons floraux; graines	CX191460
<i>BvW LIM1</i>	<i>Beta vulgaris</i>	2	Jeunes racines	CV301701
<i>BvW LIM2</i>	<i>Beta vulgaris</i>	4	Inflorescences entières; racines entières	BI543577
<i>CcW LIM1</i>	<i>Coffea canephora</i>	9	Jeunes feuilles; endosperme et péricarpe, 30 semaines après pollinisation au stade moyen du développement de la graine; péricarpe, à tous les stades de développement	DV689830
<i>CIW LIM2</i>	<i>Curcuma longa</i>	4	Jeunes feuilles	DY384040
<i>CeW LIM1</i>	<i>Cichorium endivia</i>	2	Multiple sources d'ARN	EL360541
<i>CePLIM1</i>	<i>Cichorium endivia</i>	3	Multiple sources d'ARN	EL353679
<i>CeW LIM2</i>	<i>Cichorium endivia</i>	4	Multiple sources d'ARN	EL362123
<i>CiβLIM1</i>	<i>Cichorium intybus</i>	3	Multiple sources d'ARN, feuilles au 4 <sup>e</sup> jour d'induction d'embryogenèse somatique	EH690741
<i>CiPLIM1</i>	<i>Cichorium intybus</i>	2	Multiple sources d'ARN	EH706676
<i>CiPLIM2</i>	<i>Cichorium intybus</i>	2	Multiple sources d'ARN	EH690878
<i>CesPLIM1</i>	<i>Centaurea solstitialis</i>	1	Multiple sources d'ARN	EH788526
<i>CesβLIM1a</i>	<i>Centaurea solstitialis</i>	2	Multiple sources d'ARN	EH763915
<i>CesβLIM1b</i>	<i>Centaurea solstitialis</i>	1	Multiple sources d'ARN	EH787223
<i>CesW LIM2a</i>	<i>Centaurea solstitialis</i>	1	Multiple sources d'ARN	EH770323
<i>CesW LIM2b</i>	<i>Centaurea solstitialis</i>	4	Multiple sources d'ARN	EH752499
<i>CmβLIM1a</i>	<i>Centaurea maculosa</i>	2	Multiple sources d'ARN	EH729124
<i>CmW LIM1</i>	<i>Centaurea maculosa</i>	3	Multiple sources d'ARN	EH743139
<i>CmW LIM2b</i>	<i>Centaurea maculosa</i>	2	Multiple sources d'ARN	EH742347
<i>CmPLIM2</i>	<i>Centaurea maculosa</i>	8	Multiple sources d'ARN	EH726468
<i>CsaLIM1</i>	<i>Citrus sinensis</i>	28	Jeunes plants entiers (8); sacs de jus (1); cals (2); fleurs (2); phloème (2); racines de jeunes plants infectées par des nématodes (7); racines	CK933805

## Annexes

			de jeunes plants déficientes en fer (5)	
Cs $\beta$ LIM1	<i>Citrus sinensis</i>	5	Cals (2); jeunes plants entiers (1); pulpe (1); ovaires immatures (1)	CV720040
CsWLM2	<i>Citrus sinensis</i>	38	Cals; fleurs; méristèmes; plantes entières; fruits; développement de fruits flavédo à 80 JAF; pulpe de fruits; albédo mûr; autres	CX676366
Ct $\beta$ LIM1	<i>Carthamus tinctorius</i>	1	Multiple sources d'ARN	EL407730
CtPLIM1	<i>Carthamus tinctorius</i>	3	Multiple sources d'ARN	EL410027
CtWLM1	<i>Carthamus tinctorius</i>	4	Multiple sources d'ARN	EL404658
CtWLM2	<i>Carthamus tinctorius</i>	2	Multiple sources d'ARN	EL405435
CtPLIM2	<i>Carthamus tinctorius</i>	1	Multiple sources d'ARN	EL407661
CrLIM1	<i>Ceratopteris richardii</i>	2	Spores	contig
EcWLM1	<i>Eleusine coracana</i>	1	Racines	EB086113
EeWLM1	<i>Euphorbia esula</i>	3	Tissus mélangés	contig
EeaLIM1	<i>Euphorbia esula</i>	1	Tissus mélangés	DV124183
ElgWLM2	<i>Elaeis guineensis</i>	Non trouvé	Non trouvé	<b>CPHO63</b>
Et $\beta$ LIM1	<i>Euphorbia tirucalli</i>	4	Cals induits à partir de tiges	BP962670
EeWLM2	<i>Euphorbia esula</i>	4	Tissus mélangés, cals induits à partir de tiges	DV130268
EePLIM2	<i>Euphorbia esula</i>	2	Tissus mélangés	DV140326
EgWLM1	<i>Eucalyptus globulus</i>	0	Non trouvé	<b>AB208709</b>
FvWLM2	<i>Fragaria vesca</i>	1	Jeunes plants stressé par le froid	DY672997
GhPLIM1a	<i>Gerbera hybrida</i>	10	Etamines	AJ758658
GmWLM1a	<i>Glycine max</i>	54	Graines; plantes entières; tiges; racines; feuilles; cosses; fleurs; hypocotyles de soja infectés par <i>phytophthora</i> ; hypocotyles, jeunes plants étiolés âgés de 9-10 jour; jeunes plants entiers; tiges après 24 heures de germination; mélange; autres	contig
GmWLM1b	<i>Glycine max</i>	12	Jeunes plants	contig
Gm $\beta$ LIM1	<i>Glycine max</i>	14	racines déficientes en eau (6); racines (1); graines (6); tiges (1)	CX709969
GmWLM2	<i>Glycine max</i>	30	racines déficientes en eau (10); tiges (6); graines (6); fleurs (2); plantes entières (2); autres (3); inconnu (1)	CX703844
GhWLM1	<i>Gossypium hirsutum</i>	7	Fibre, 5-10 jours post-anthèse JPA (1); cotylédons (1); jeunes plants entiers (4); fleur (1)	DR454855
GhaLIM1	<i>Gossypium hirsutum</i>	33	Fibres de coton au stade de formation de la paroi secondaire, 20 JPA (29); fibres 5-10 JPA (2); jeunes plants entiers (2)	contig
Ga $\beta$ LIM1	<i>Gossypium arboreum</i>	2	Fibres isolées de graines récoltées 7-10 JPA; fibres immatures, six JPA	BQ409025
GhWLM2a (GhLIM-1)	<i>Gossypium hirsutum</i>	7	Ovules de coton et fibres; fibres 7-10 JPA	<b>AF443117</b>
GaWLM2b	<i>Gossypium arboreum</i>	14	Fibres isolées de graines récoltées 7-10 JPA (8); fibres immatures, six jours JPA (1); fibres à 5-10 JPA (1); zones d'abscission des graines de coton (1); fibres et ovules à -3 à 25 JPA (1); graines de coton à 8-10 JPA (1); jeunes plants (1)	BM359147
GrPLIM2a	<i>Gossypium raimondii</i>	5	Organes floraux -3 à +3 JPA; fibres immatures à 6 JPA; fibres à 5-10 JPA	contig
GrPLIM2b	<i>Gossypium raimondii</i>	1	Organes floraux -3 à +3 JPA	CO122392
GrPLIM2c	<i>Gossypium raimondii</i>	2	Organes floraux -3 à +3 JPA	CO128788
HaWLM1	<i>Helianthus annuus</i>	7	Multiple sources d'ARN	<b>AF187105</b>
HaPLIM1a-SF3	<i>Helianthus annuus</i>	0	Non trouvé	<b>AF187104</b>
HaPLIM1b	<i>Helianthus annuus</i>	23	Multiple sources d'ARN	<b>AF187103</b>
Ha $\beta$ LIM1	<i>Helianthus annuus</i>	9	Multiple sources d'ARN, protoplastes âgés de 1 à 5 jours	DY917845
H $\beta$ WLM2	<i>Helianthus petiolaris</i>	2	Multiple sources d'ARN, bourgeons terminaux	DY950080
HaPLIM2	<i>Helianthus annuus</i>	49	Multiple sources d'ARN	<b>AF047353</b>
Ha $\delta$ LIM2b	<i>Helianthus annuus</i>	2	Multiple sources d'ARN	DY914658
Ha $\delta$ LIM2a	<i>Helianthus annuus</i>	1	Multiple sources d'ARN	DY916043
HvWLM1	<i>Hordeum vulgare</i>	44	Tous tissus incluant les inflorescences, les tiges, les racines, les épis avant anthèse et l'embryon 14 JPA	contig
HvWLM2	<i>Hordeum vulgare</i>	89	Tous tissus incluant les inflorescences, les racines, les épis, les carpelles, le sac embryonnaire à 0-10 JPA	CK125568
HvPLIM2	<i>Hordeum vulgare</i>	3	Anthères au stade jaune	contig
InWLM1b	<i>Ipomoea nil</i>	2	Mixture de fleurs et de bourgeons floraux	contig
InWLM1a	<i>Ipomoea nil</i>	3	Mixture de fleurs et de bourgeons floraux	BJ569751
In $\beta$ LIM1	<i>Ipomoea nil</i>	12	Mixture de fleurs et de bourgeons floraux (6); tiges de jeunes plants âgés de 8 jours (4); feuilles (1)	BJ576632
InWLM2	<i>Ipomoea nil</i>	10		CJ773251
LeWLM1	<i>Lycopersicon esculentum</i>	6	Fruits; fleurs; fruits matures; racines	BW685280
LpPLIM1	<i>Lycopersicon pennellii</i>	26	Pollen exclusivement	contig
Le $\beta$ LIM1a	<i>Lycopersicon esculentum</i>	7	Racines (4); cultures en suspension (1); fleurs (2)	<b>AY796114</b>
Le $\beta$ LIM1b	<i>Lycopersicon esculentum</i>	9	Fruits; racines; feuilles; cultures en suspension; fleurs; bourgeons de 3-8 mm	BW686491
LeWLM2	<i>Lycopersicon esculentum</i>	21	Racines (1); cals (6); épiderme (1); fleurs (5); feuilles (3); tiges (1); fruits (3); plantes entières (1)	CN385862
Lp $\delta$ LIM2	<i>Lycopersicon pennellii</i>	4	Pollen	contig

## Annexes

<i>LsWLM1</i>	<i>Lactuca sativa</i>	8	Multiple sources d'ARN	DW131741
<i>LsPLIM1</i>	<i>Lactuca saligna</i>	3	Multiple sources d'ARN	DW049592
<i>LsβLIM1</i>	<i>Lactuca sativa</i>	12	Multiple sources d'ARN	DY977036
<i>LsWLM2a</i>	<i>Lactuca sativa</i>	2	Multiple sources d'ARN	DY981728
<i>LsWLM2b</i>	<i>Lactuca sativa</i>	19	Multiple sources d'ARN	DW140776
<i>MdWLM1</i>	<i>Malus x domestica</i>	9	Xylème, bois mature (2); fruits (5); bourgeons végétatifs pré ouvrant (1); fleurs (1)	EB138908
<i>MdWLM2</i>	<i>Malus x domestica</i>	36	Feuilles; entrenoeuds de tiges; bourgeons végétatifs; bouts de racines	contig
<i>MdδLIM2</i>	<i>Malus x domestica</i>	3	Fleurs	contig
<i>MeWLM1</i>	<i>Manihot esculenta</i>	8	Tiges (3); racines (5)	contig
<i>MeWLM2</i>	<i>Manihot esculenta</i>	2	Jeunes feuilles et méristèmes apicaux; feuilles matures et pétiole; racines; tubercules et pelures de tubercules de plantes en stress hydrique	DV452494
<i>MgWLM2</i>	<i>Mimulus guttatus</i>	3	Bouts d'inflorescences se développant, consistant en de très jeunes bourgeons floraux non ouverts (< 5mm de long), méristèmes floraux et méristèmes apicaux de tiges	contig
<i>MtWLM1</i>	<i>Medicago truncatula</i>	14	Cosses incluant les graines (1); racines (6); mycorhizes arbusculaires (1); fleurs (2); feuilles (3); tiges (1)	AJ498553
<i>MtWLM2</i>	<i>Medicago truncatula</i>	28	Racines; feuilles; fleurs; jeunes plants; mycorhizes arbusculaires; nodules de racines symbiotiques; plantules; graines en germination; tiges infectées de puceron; tissus mélangés	BG585852
<i>McWLM2</i>	<i>Mesembryanthemum crystallinum</i>	9	Feuilles (8); méristèmes apicaux et primordium de feuilles (1)	contig
<i>NtWLM1</i>	<i>Nicotiana tabacum</i>	2	Tissus mélangés	<b>AF184109, AB079513</b>
<i>NtLIM1</i>	<i>Nicotiana tabacum</i>	2	Tissus mélangés	<b>AB023479</b>
<i>NtPLIM1a</i>	<i>Nicotiana tabacum</i>	0	Non trouvé	<b>AF197567</b>
<i>NtPLIM1b</i>	<i>Nicotiana tabacum</i>	0	Non trouvé	<b>AF197568</b>
<i>NtβLIM1a</i>	<i>Nicotiana tabacum</i>	4	Feuilles (1); feuilles et tiges (2); BY2 (1)	DV999848
<i>NtβLIM1b</i>	<i>Nicotiana tabacum</i>	5	Racines (2); feuilles et tiges (2); BY2 (1)	DW004416
<i>NtWLM2</i>	<i>Nicotiana tabacum</i>	3	Fleurs; feuilles et tiges; BY2	<b>Y11002</b>
<i>NtPLIM2</i>	<i>Nicotiana tabacum</i>	3	Fleurs	<b>AF116851</b>
<i>NtδLIM2</i>	<i>Nicotiana tabacum</i>	2	Fleurs	EB426388
<i>ObWLM2</i>	<i>Ocimum basilicum</i>	2	Trichomes glandulaires	DY333831
<i>PaWLM2</i>	<i>Prunus armeniaca</i>	15	Fruits au stade vert; fruits au stade mûr; mésocarpe de fruits (11); mésocarpe de fruits avec épiderme; feuilles	CV047143
<i>PgLIM1a</i>	<i>Picea glauca</i>	27	Systèmes racinaires entiers (7); cambium et phloème (9); xylème en différenciation non lignifié (5); xylème en différenciation (1); aiguilles (1); strobiles (2); tissus divers (2)	DV986756
<i>PgLIM1b</i>	<i>Picea glauca</i>	29	Phloème (3); cambium et phloème (8); cônes entiers (4); racines (5); xylème (2); écorce avec le phloème et le cambium attaché (2); parties vertes des tissus de l'apex (3); tissus variés (2)	DR563806
<i>PgWLM2</i>	<i>Picea glauca</i>	47	Racines (7); phloème (9); xylème secondaire (2); bourgeons de tiges jeunes et matures (5); cônes mâles (4); tissus d'embryon somatique (2); écorce avec le phloème et le cambium attaché (14); parties vertes des tissus de l'apex (4)	CO234567
<i>PGPS/D1</i>	<i>Petunia x hybrida</i>	2	Tous les organes floraux	<b>AF049917</b>
<i>PhWLM2</i>	<i>Petunia x hybrida</i>	5	Tous les organes floraux	contig
<i>PitLIM1a</i>	<i>Pinus taeda</i>	62	Racines (42); xylème (14); cultures en suspension d'embryons somatiques (2); bouts de tiges (3); cônes de pollen (1)	CF478208
<i>PitLIM1b</i>	<i>Pinus taeda</i>	33	Xylème (12); racines (15); bouts de tiges (2); inconnu (4)	CV035130
<i>PitWLM2</i>	<i>Pinus taeda</i>	32	Racines (14); cultures en suspension d'embryons somatiques (5); bouts de tiges (4); xylème (7); inconnu (2)	DR695125
<i>PpLIM</i>	<i>Physcomitrella patens</i>	12	Protoplastes (chloronémata) (9); mélange de chloronémata et jeunes gamétophores avec 2 à 5 feuilles (2); mélange de chloronémata, caulonémata et protonémata semblable aux rhizoïdes (1)	BJ946868
<i>PvWLM1</i>	<i>Panicum virgatum</i>	1	Entrenoeuds	DN143962
<i>PvWLM2</i>	<i>Panicum virgatum</i>	2	Epis; entrenoeuds	DN140706
<i>RcaLIM1</i>	<i>Ricinus communis</i>	1	Racines	EG667055
<i>RcβLIM1</i>	<i>Ricinus communis</i>	1	Graines	EG660094
<i>RcWLM2</i>	<i>Ricinus communis</i>	2	Feuilles	EE257544
<i>RcPLIM2</i>	<i>Ricinus communis</i>	3	Feuilles	EE255361
<i>SeaPLIM1</i>	<i>Senecio aethnensis</i>	6	Fleurs et bourgeons floraux sans la bractée à 1-4 jours pré anthèse	CO553805
<i>SecδLIM2</i>	<i>Senecio chrysanthemifolius</i>	4	Fleurs et bourgeons floraux sans la bractée à 1-4 jours pré anthèse	DY665121
<i>SbWLM2</i>	<i>Sorghum bicolor</i>	16	Racines traitées par de l'acide et des bases (4); jeunes plantes déficientes en phosphore (4); feuilles et racines avec stress oxydatif (1); jeunes plants poussant dans l'obscurité (6); jeunes plants poussant à la lumière (1)	contig
<i>ShaLIM1</i>	<i>Saruma henryi</i>	4	Bourgeons floraux de 2,5 mm	DT601849
<i>ShWLM2a</i>	<i>Saruma henryi</i>	3	Bourgeons floraux de 2,5 mm	contig
<i>ShWLM2b</i>	<i>Saruma henryi</i>	2	Bourgeons floraux de 2,5 mm	contig
<i>SoWLM1a</i>	<i>Saccharum officinarum</i>	54	Zones de transition tige-racine de jeunes plantes et plantes adultes	CA152933

## Annexes

			(12); feuilles enroulée incluant l'apex (4); bourgeons latéraux (2); bouts de racines (3); graines en développement (1); écorce de tiges de plantes adultes (4); méristèmes apicaux et tissus l'entourant de plantes matures (10); inflorescences au début de développement (1cm de long) (9); premiers entrenoeuds apicaux de tiges de plantes adultes (1); bases d'inflorescences en développement (5 cm de long) (3); Inflorescences développées et rachis (20 cm de long) (5)	
<i>SoWLM1b</i>	<i>Saccharum officinarum</i>	75	Zones de transition tige-racine de plantes adultes (74), méristèmes apicaux et tissus l'entourant de plantes matures (1)	CA154735
<i>SoWLM2</i>	<i>Saccharum officinarum</i>	20	Feuilles étiolées de jeunes plants cultivés in vitro (1); premiers et quatrièmes entrenoeuds apicaux de tiges de plantes adultes (3); jeunes plants inoculés avec <i>Herbaspirillum rubrisubalbicans</i> (1); inflorescences développée et rachis (20 cm de long) (1); inflorescences au début de développement (1cm de long) (1); ensemble de cals de canne à sucre soumis à des stress de température (4 C et 37 C) (1); méristèmes apicaux et tissus l'entourant de plantes immatures (1); écorce de tiges de plantes adultes (1); zones de transition tige-racine de jeunes plantes (1); apex de racines de plantes adultes (4); bourgeons latéraux de plantes adultes (1); feuilles (4)	CA299778
<i>StWLM1</i>	<i>Solanum tuberosum</i>	4	Tissus mélangés (2); tubercules en développement (1); petites feuilles et pétioles (1)	contig
<i>ScβLIM1a</i>	<i>Solanum chacoense</i>	2	Embryons zygotiques à globulaires 2-6 jours après pollinisation des ovaires	contig
<i>StβLIM1b</i>	<i>Solanum tuberosum</i>	8	Tubercules en développement (3); feuilles infectées avec <i>P. infestans</i> (2); cultures en suspension dérivées de cals (1); tissus mélangés (2)	contig
<i>StWLM2</i>	<i>Solanum tuberosum</i>	27	Embryons floraux (1); tissus mélangés (2); tubercules (4); feuilles (4); racines (3); stolons (3); tubercules en germination (1); ovaires 2-6 jours post-pollinisation (1); petites feuilles et pétioles (5); bourgeons axillaires d'explants de tiges, stolons gonflant (3)	contig
<i>TaWLM1</i>	<i>Triticum aestivum</i>	48	Racines; graines; ovaires; embryon germant; inflorescences; épis; fleurs; anthères; couronnes et feuilles; plantules	CV762492
<i>TaWLM2</i>	<i>Triticum aestivum</i>	134	Inflorescences (épis et épillet); couronnes; feuilles; racines; graines; cals; anthères	CV777980
<i>TaPLIM2a</i>	<i>Triticum aestivum</i>	14	Epis avant anthèse (12); épis a la date de floraison (1); anthères (1)	contig
<i>TaPLIM2b</i>	<i>Triticum aestivum</i>	11	Epis avant anthèse (2); épis a la date de floraison (7); têtes à l'anthèse (2)	contig
<i>TaPLIM2c</i>	<i>Triticum aestivum</i>	12	Epis avant anthèse (2); épis a la date de floraison (2); têtes à l'anthèse (1); anthères (2); racines	contig
<i>ToPLIM1</i>	<i>Taraxacum officinale</i>	2	Multiple sources d'ARN	DY841300
<i>TkβLIM1</i>	<i>Taraxacum kok-saghyz</i>	6	Multiple sources d'ARN (2), racines (4)	contig
<i>ToWLM2</i>	<i>Taraxacum officinale</i>	10	Multiple sources d'ARN (8), racines (2)	DY811951
<i>ToδLIM2</i>	<i>Taraxacum officinale</i>	1	Multiple sources d'ARN	DY818558
<i>TrLIM</i>	<i>Tortula ruralis</i>	3	Gamétophytes réhydratés	CN207416
<i>VvaLIM1</i>	<i>Vitis vinifera</i>	5	Tiges (2); pétioles (3)	CF211891
<i>VvβLIM1</i>	<i>Vitis vinifera</i>	17	Fruits au stade de mûrissement (1); baies 8 à 16 semaine JAF (4); baies (4); tiges (2); embryon (1); jeunes racines (1); cultures de suspension de cellules (1); fruits sans graines (2); raisins (1)	CF213597
<i>VvWLM2</i>	<i>Vitis vinifera</i>	36	Inflorescences incluant les fleurs (10); fruits de la baie (16); fruits sans graines (2); bourgeons (3); péricarpes du fruit (2); fleurs (2); feuilles (1)	CB974935
<i>VvPLIM2a</i>	<i>Vitis vinifera</i>	2	Fleurs, feuilles et racines	EC955522
<i>ZoβLIM1</i>	<i>Zingiber officinale</i>	7	Rhizomes (6); racines (1)	contig
<i>ZmWLM1</i>	<i>Zea mays</i>	194	Epis; fleurs; méristèmes; péricarpes; racines; méristèmes apicaux de tiges végétatives et primordium de feuilles au stade P1-P4; tissus mélangés; épis immatures; primordium de feuilles immatures et méristèmes végétatifs; jeunes plants et soies; glands; sac embryonnaire; plantes entières; jeunes plants sans grain; gamétophytes femelles;	<b>DR826954, AY112454</b>
<i>ZmWLM2</i>	<i>Zea mays</i>	86	Tissus mélangés; racines; méristèmes apicaux végétatif et les primordiums de feuilles au stade P1-P4; grains en développement; jeunes plants sans grain; ovules non fertilisées; glands; jeunes tiges végétatives; jeunes plants et soies; stades mélangés d'anthères et de pollen; épis; primordium de feuilles; feuilles d'épis; apex; nucelle, embryons et endosperme; feuilles et tiges; embryons; épis non matures	<b>AY103686</b>
<i>ZmPLIM2a</i>	<i>Zea mays</i>	201	Pollen (87); pollen mature (66); mélange de stades d'anthères et de pollen (3); tissus mélangés ou non caractérisés (45)	EB167114
<i>ZmPLIM2b</i>	<i>Zea mays</i>	14	Pollen (4); tissus mélangés ou non caractérisés (10)	DV169657



## Annexes

### Annexe 3. Classification phylogénétique et taxonomie des espèces de plantes pour lesquelles des gènes *LIM* ont été identifiées (The Angiosperm Phylogeny Group, 2003). (ND = Non définie)

Espèce	Nom commun	Genre	Tribu	Sous-famille	Famille	Ordre	Sous-classe	Classe	Division
<i>Acorus americanus</i> (Raf.)	Acore commun ou Roseau aromatique	Acorus	ND	ND	Acoraceae	Acorales	ND	Liliopsida	Magnoliophyta
<i>Aquilegia formosa</i> x <i>A. pubescens</i>	Ancolies	Aquilegia	ND	ND	Ranunculaceae	Ranunculales	ND	Eudicots (Magnoliopsida)	Magnoliophyta
<i>Antirrhinum majus</i>	Muflier ou gueule de loup	Antirrhinum	ND	ND	Plantaginaceae	Lamiales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Arabidopsis thaliana</i>	Arabette	Arabidopsis	ND	ND	Brassicaceae (ou crucifères)	Brassicales	rosids (eurosids II)	Core Eudicots	Magnoliophyta
<i>Brachypodium distachyon</i>	Brachypode à deux épillets	Brachypodium	Brachypodieae	Pooideae	Poaceae (BEP clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Brassica napus</i>	Colza	Brassica	ND	ND	Brassicaceae (ou crucifères)	Brassicales	rosids (eurosids II)	Core Eudicots	Magnoliophyta
<i>Beta vulgaris</i>	Betterave	Beta	ND	ND	Amaranthaceae	Caryophyllales	ND	Core Eudicots	Magnoliophyta
<i>Coffea canephora</i>	Caféier robusta	Coffea	Coffeae	Ixoroideae	Rubiaceae	Gentianales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Curcuma longa</i>	Curcuma long	Curcuma	ND	ND	Zingiberaceae	Zingiberales	commelinids	Liliopsida	Magnoliophyta
<i>Cichorium endivia</i>	Chicorée endive (frisée, scarole...)	Cichorium	Cichorieae	Cichorioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Cichorium intybus</i>	Chicorée sauvage ou endive	Cichorium	Cichorieae	Cichorioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Centaurea solstitialis</i>	Centaurée du solstice	Centaurea	Cardueae	Carduoideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Centaurea maculosa</i>	Centaurée maculée	Centaurea	Cardueae	Carduoideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Citrus sinensis</i>	Oranger	Citrus	ND	ND	Rutaceae	Sapindales	rosids (eurosids II)	Core Eudicots	Magnoliophyta
<i>Carthamus tinctorius</i>	Carthame des teinturiers	Carthamus	Cardueae	Carduoideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Ceratopteris richardii</i>	Fougère d'eau en triangle	Ceratopteris	ND	ND	Pteridaceae	Filicales	ND	Filicopsida	(Pteridophyta)
<i>Eleusine coracana</i>	Mil africain, éleusine	Eleusine	Eragrostideae	Chloridoideae	Poaceae (PACCAD clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Euphorbia esula</i>	Euphorbe érule ou âcre	Euphorbia	Euphorbieae	Euphorbioideae	Euphorbiaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Elaeis guineensis</i>	Palmier à huile d'Afrique	Elaeis	Cocoseae (Elaeidinae)	Arecoideae	Arecaceae	Arecales	commelinids	Liliopsida	Magnoliophyta
<i>Euphorbia tirucalli</i>	Euphorbe crayon ou effilée	Euphorbia	Euphorbieae	Euphorbioideae	Euphorbiaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Eucalyptus globulus</i>	Eucalyptus ou gommier bleu	Eucalyptus	ND	ND	Myrtaceae	Myrtales	rosids	Core Eudicots	Magnoliophyta
<i>Fragaria vesca</i>	Fraisier	Fragaria	ND	Rosoideae	Rosaceae	Rosales	rosids (eurosids I)	Core Eudicots	Magnoliophyta

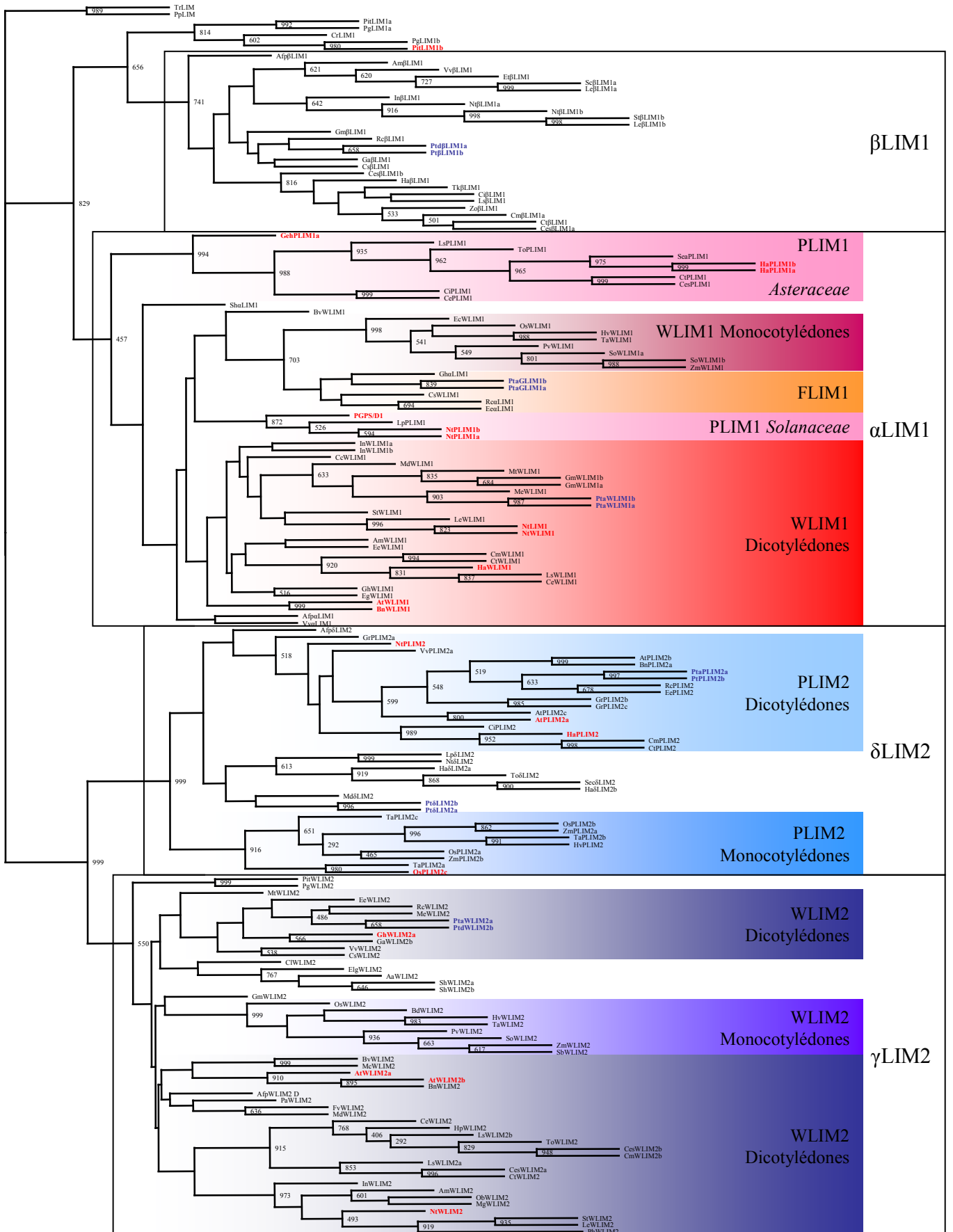
## Annexes

<i>Gerbera hybrida</i>	Gerbera hybride	Gerbera	Mutisieae	Mutisioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Glycine max</i>	Soja	Glycine	Phaseoleae	Papilionoideae	Fabaceae	Fabales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Gossypium arboreum</i>	Cotonnier en arbre	Gossypium	ND	Malvoideae	Malvaceae	Malvales	rosids (eurosids II)	Core Eudicots	Magnoliophyta
<i>Gossypium hirsutum</i>	Cotonnier	Gossypium	ND	Malvoideae	Malvaceae	Malvales	rosids (eurosids II)	Core Eudicots	Magnoliophyta
<i>Gossypium raimondii</i>	Cotonnier	Gossypium	ND	Malvoideae	Malvaceae	Malvales	rosids (eurosids II)	Core Eudicots	Magnoliophyta
<i>Helianthus annuus</i>	Tournesol	Helianthus	Heliantheae	Asteroideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Helianthus petiolaris</i>	Tournesol	Helianthus	Heliantheae	Asteroideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Hordeum vulgare</i>	Orge	Hordeum	Triticeae	Pooideae	Poaceae (BEP clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Ipomoea nil</i>	Ipomée	Ipomoea	Ipomoeae	ND	Convolvulaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Lycopersicon esculentum</i>	Tomate	Solanum (Lycopersicon)	ND	Solanoideae (Solaneae)	Solanaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Lycopersicon pennellii</i>	Tomate	Solanum (Lycopersicon)	ND	Solanoideae (Solaneae)	Solanaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Lactuca saligna</i>	Laitue à feuille de saule	Lactuca	Cichorieae	Cichorioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Lactuca sativa</i>	Laitue cultivée	Lactuca	Cichorieae	Cichorioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Malus x domestica</i>	Pommier	Malus	ND	Maloideae	Rosaceae	Rosales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Manihot esculenta</i>	Manioc	Manihot	Manihoteae	Crotonoideae	Euphorbiaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Mimulus guttatus</i>	Mimule tacheté	Mimulus	ND	ND	Phrymaceae	Lamiales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Medicago truncatula</i>	Luzernes	Medicago	Trifolieae	Papilionoideae	Fabaceae	Fabales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Mesembryanthemum crystallinum</i>	Mésembryanthème à cristaux	Mesembryanthemum	ND	ND	Aizoaceae	Caryophyllales	ND	Core Eudicots	Magnoliophyta
<i>Nicotiana tabacum</i>	Tabac	Nicotiana	ND	Nicotianoideae (Nicotianeae)	Solanaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Ocimum basilicum</i>	Basilic	Ocimum	Ocimeae	Nepetoideae	Lamiaceae	Lamiales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Oryza sativa</i>	riz	Oryza	Oryzeae	Ehrhartoideae	Poaceae (BEP clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Prunus armeniaca</i>	Abricotier	Prunus	ND	Amygdaloideae	Rosaceae	Rosales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Picea glauca</i>	Épinette blanche ou épicéa	Picea	ND	ND	Pinaceae	Coniferales	ND	Coniferopsida	Coniferophyta

## Annexes

<i>Petunia x hybrida</i>	Pétunia	Petunia	ND	Petunioideae	Solanaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Pinus taeda</i>	Pin à l'encens ou à torches	Pinus (Pinus)	ND	ND	Pinaceae	Coniferales	ND	Coniferopsida	Coniferophyta
<i>Physcomitrella patens</i>	Mousse	Physcomitrella	ND	ND	Funariaceae	Funariales	Funariidae	Bryopsida	Bryophyta (Moss Superclass V)
<i>Panicum virgatum</i>	Panic dressé ou panic érigé	Panicum	Paniceae	Panicoideae	Poaceae (PACCAD clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Populus nigra</i>	Peuplier noir	Populus	Saliceae	ND	Salicaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Populus tremula x P. alba</i>	Peuplier tremble x P. blanc	Populus	Saliceae	ND	Salicaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Populus trichocarpa</i>	Peuplier baumier	Populus	Saliceae	ND	Salicaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Ricinus communis</i>	Ricin	Ricinus	Acalypheae	Acalyphoideae	Euphorbiaceae	Malpighiales	rosids (eurosids I)	Core Eudicots	Magnoliophyta
<i>Senecio aethnensis</i>	Séneçon	Senecio	Senecioneae	Asteroideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Senecio chrysanthemifolius</i>	Séneçon	Senecio	Senecioneae	Asteroideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Sorghum bicolor</i>	Sorgho	Sorghum	Andropogoneae	Panicoideae	Poaceae (PACCAD clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Saruma henryi</i>	ND	Saruma	ND	ND	Aristolochiaceae	Piperales	magnoliids	ND	Magnoliophyta
<i>Saccharum officinarum</i>	Canne à sucre	Saccharum	Andropogoneae	Panicoideae	Poaceae (PACCAD clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Solanum tuberosum</i>	Pomme de terre	Solanum	ND	Solanoideae (Solaneae)	Solanaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Solanum chacoense</i>	Pomme de terre	Solanum	ND	Solanoideae (Solaneae)	Solanaceae	Solanales	asterids (lamiids)	Core Eudicots	Magnoliophyta
<i>Triticum aestivum</i>	Blé tendre, ou froment	Triticum	Triticeae	Pooideae	Poaceae (BEP clade)	Poales	commelinids	Liliopsida	Magnoliophyta
<i>Taraxacum officinale</i>	Pissenlit commun	Taraxacum	Cichorieae	Cichorioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Taraxacum kok-saghyz</i>	Pissenlit	Taraxacum	Cichorieae	Cichorioideae	Asteraceae	Asterales	asterids (campanulids)	Core Eudicots	Magnoliophyta
<i>Tortula ruralis</i>	Mousse	Tortula	ND	ND	Pottiaceae	Pottiales	Dicranidae	Bryopsida	Bryophyta (Moss Superclass V)
<i>Vitis vinifera</i>	Vigne	Vitis	ND	ND	Vitaceae	Vitales	ND	Core Eudicots	Magnoliophyta
<i>Zingiber officinale</i>	Gingembre	Zingiber	ND	ND	Zingiberaceae	Zingiberales	commelinids	Liliopsida	Magnoliophyta
<i>Zea mays</i>	Maïs	Zea	Andropogoneae	Panicoideae	Poaceae (PACCAD clade)	Poales	commelinids	Liliopsida	Magnoliophyta

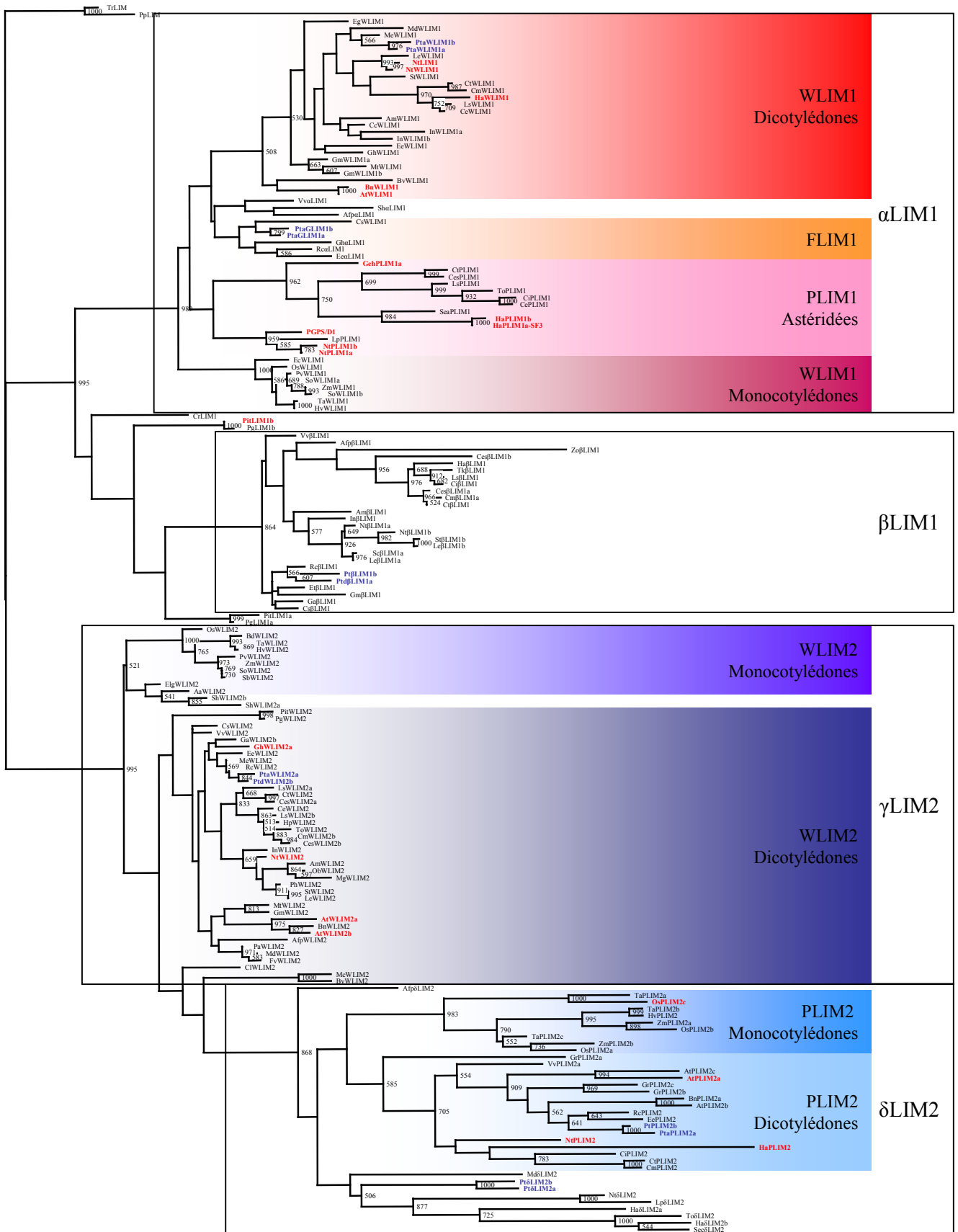
Annexe 4. Voir légende à la page suivante



### **Annexe 4.** Arbre phylogénétique de la famille de protéines à domaine LIM chez les plantes.

Les séquences d'acides aminés de 178 protéines à domaine LIM ont été alignées au moyen du logiciel ClustalW, puis analysées avec la méthode de parcimonie en utilisant le programme PHYLIP. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. Les protéines à domaine LIM ont été renommées en suivant leur appartenance aux groupes ou sous-groupes WLIM1, PLIM1, FLIM1,  $\alpha$ LIM1,  $\beta$ LIM1, WLIM2, PLIM2 ou  $\delta$ LIM2 qui sont encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

Annexe 5. Voir légende à la page suivante



### **Annexe 5.** Arbre phylogénétique de la famille de protéines à domaine LIM chez les plantes.

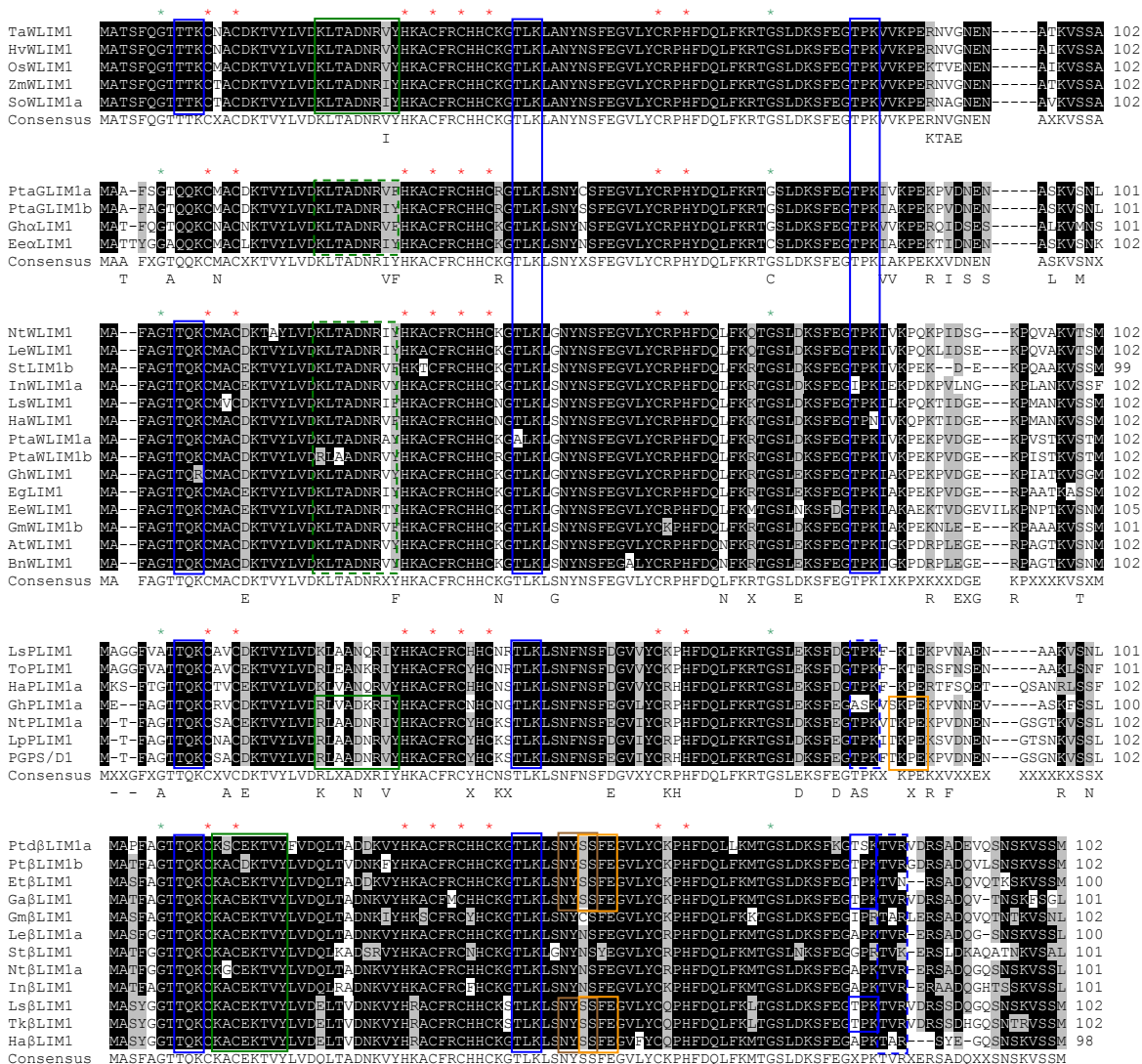
Les séquences d'acides aminés de 178 protéines à domaine LIM ont été alignées au moyen du logiciel ClustalW, puis analysées avec la méthode de maximum de vraisemblance en utilisant les programmes PhymI et BIONJ. Les nombres situés au niveau des noeuds représentent des valeurs de bootstrap ( $\geq 500$ ) fondées sur 1000 répétitions. La longueur des branches est proportionnelle aux nombres attendus de substitutions en acides aminés par site, et une échelle est indiquée en dessous de l'arbre. Les protéines à domaine LIM ont été renommées en suivant leur appartenance aux groupes ou sous-groupes WLIM1, PLIM1, FLIM1,  $\alpha$ LIM1,  $\beta$ LIM1, WLIM2, PLIM2 ou  $\delta$ LIM2 qui sont encadrés. Les protéines à domaine LIM de peuplier sont marquées en bleu gras, et les protéines à domaine LIM de plante pour lesquelles le profil d'expression a été déterminé sont marquées en rouge gras. Comme indiqué dans la figure 39, un acronyme d'espèce est ajouté avant chaque nom de protéine à domaine LIM. Pour des détails additionnels sur chaque gène, le lecteur pourra se référer à l'annexe 2.

## Annexes

### Annexe 6. Alignement des séquences protéiques des domaines LIM1 et LIM2

L'alignement de séquence a été réalisé à partir d'une sélection de protéines à domaine LIM, et les domaines N-terminal, LIM1 et inter-LIM ont été séparés des domaines LIM2 et C-terminal. Les protéines orthologues LIM ont été groupées en fonction de leurs appartenances aux différents groupes et sous-groupes LIM. Le domaine C-terminal des protéines PLIM2 a été coupé par manque de place sur la figure. Les Gaps indiqués par des tirets ont été gardés ou ajoutés pour souligner les motifs communs entre les sous-groupes. L'ombrage des acides aminés a été réalisé avec le logiciel Bioedit en utilisant un seuil de 90 % pour les résidus identiques et les résidus similaires (déterminés par la matrice Blosum). Ces résidus sont respectivement représentés par un ombrage noir et gris. Les séquences consensus sont indiquées en dessous de chaque alignement et ont été déterminées avec un seuil de 90 % pour chaque groupe et sous-groupe (X = n'importe quel acide aminé). Les astérisques rouges au dessus des alignements marquent les ligands des doigts de zinc. Les astérisques oranges montrent le ligand histidine alternatif du second doigt de zinc du deuxième domaine LIM. Les astérisques verts montrent les glycines conservées entourant chaque domaine LIM. Les boîtes brunes soulignent les sites putatifs de glycosylation sur les asparagines (ASN). Les boîtes vertes, oranges, bleues et violettes indiquent respectivement les sites putatifs de phosphorylation sur les tyrosines (Tyr), et les sites putatifs de phosphorylation par les caséines kinase (CKII), les protéines kinase C (PKC), et les protéines kinase dépendantes du cAMP et du cGMP (cAMP) identifiés par les analyses Prosite et MEME. Les sites de glycosylation ou de phosphorylation non conservés sont indiqués par des boîtes hachurées.

#### Domaines N-terminal, LIM1 et inter-LIM de protéines des groupes $\alpha$ LIM1 et $\beta$ LIM1





## Annexes

TYG X E X V D X R X S N X X R A K-D X X T X T L

### Domaines N-terminal, LIM1 et inter-LIM des protéines des groupes $\gamma$ LIM2 et $\delta$ LIM2

```
TaWLim2 M FSGTQQKCKVCTKTVPMDQLSTDGVFHRSCFKCHCKSTLSLGNYSSEFVGPYCKPHEQLFKETGSYNKSFSQSPAKSATPKLPELTRSPSKAAGM 102
HvWLim2 M FSGTQQKCKVCTKTVPMDQLSTDGVFHRSCFKCHCKSTLSLGNYSSEFVGPYCKPHEQLFKETGSYNKSFSQSPAKSATPKLPELTRSPSKAAGM 102
SoWLim2 M FSGTQQKCKVCTKTVPMDQLSTDGVFHRSCFKCHCKSTLSLGNYSSEFVGPYCKPHEQLFKETGSYNKSFSQSPAKITPELPELTRSPSKAAGM 102
ZmWLim2 M FSGTQQKCKVCTKTVPMDQLSTDGVFHRSCFKCHCKSTLSLGNYSSEFVGPYCKPHEQLFKETGSYNKSFSQSPAKITPELPELTRSPSKAAGM 102
OsWLim2 M FSGTQQKCKVCTKTVPMDQLSTDGVFHRSCFKCHCKSTLSLGNYSSEFVGPYCKPHEQLFKETGSYNKSFSQSPAKITPELPELTRSPSKAAGM 100
Consensus M FSGTQQKCKVCTKTVPMDQLSTDGVFHRSCFKCHCKSTLSLGNYSSEFVGPYCKPHEQLFKETGSYNKSFSQSPAKXAXEKLTPELTRSPSKAARM
```

A A H FCS II X A -- T D A G

```
StWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FVKS-AEKLTPELTRSPSKAAGM 100
LeWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FVKS-AEKLTPELTRSPSKAAGM 100
PhWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-AEKLTPELTRSPSKAAGM 100
NtWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-AEKLTPELTRSPSKAAGM 100
HnWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-ADKLTPELTRSPSKAAGM 100
HvWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKA-AEKLTPELTRSPSKAAGM 100
ToWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKT-VDKSSPELTRSPSKAAGM 100
LsWLim2a MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKF-AEKLTPELTRSPSKAAGM 100
AtWLim2b MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-ADKSTPELTRSPSKAAGM 100
AtWLim2a MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKPLTDKPTPELTRSPSKAAGM 101
GmWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKL-ADKTPPELTRSPSKAAGM 100
PtaWLim2a MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKT-AEKLTPELTRSPSKAAGM 100
PtdWLim2b MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-AEKLTPELTRSPSKAAGM 100
EaWLim2 MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-AEKLTPELTRSPSKAAGM 100
GmWLim2b MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS--FAKS-AEKLTPELTRSPSKAAGM 100
Consensus MSFTGTQOKKACBKTVPVELLSADGHSYHKSCFKCHCKGTLRLSNYSSEMGVLYCKPHEQLFKESGNEFNKFNFS FAKS-AEKLTPELTRSPSKAAGM
```

I G M V A R SR Q S V D Y T S S V XLXDRXX X KT RX S

```
PtaPLIM2a MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KP-ERTHD-LSRIPSKLSSV 97
PtPLIM2b MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KP-ERTHE-LIRIPSKLSSV 97
EePLIM2 MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KP-EKONE-VNRTIPSKLSSM 97
AtPLIM2b MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KT-EKP-E-LTRIPSKLSSI 96
BnPLIM2 MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KT-EKP-E-OSKIPSKLSSI 96
AtPLIM2a MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KT-EKPNHDLTRIPSKLSSV 98
GrPLIM2b MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KT-EKOND-TSRTIPSKLSSI 97
NtPLIM2 MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KA-ERON-SLTRAPSKLSSM 97
HaPLIM2 MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS--KL-NRDN-STALIPSKLSSV 97
Consensus MAFTGTLDKCAADKTVPVDMMLSEGVYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKESGNEFNKFNFS KX EXXXXXXXRTPSKXSS
```

S XV HFM LLTXD MT X R T X DY X R KX N X

```
TaPLIM2a MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 96
OsPLIM2c MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NS--QAKIPFNKRYGSV 96
TaPLIM2b MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 98
HvPLIM2 MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NS--QAKIPFNKRYGSV 98
ZmPLIM2a MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 98
OsPLIM2b MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NS--QAKIPFNKRYGSV 97
Consensus MTFVGTQDKKACDKTVHFDIDLLTADGIPYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV
```

T X E I X R K K X A N Q K X XXAV

```
Pt $\delta$ LIM2a MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 96
Pt $\delta$ LIM2b MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NS--QAKIPFNKRYGSV 96
To $\delta$ LIM2 MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 93
Ha $\delta$ LIM2b MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 93
Nt $\delta$ LIM2 MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 95
Lp $\delta$ LIM2 MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV 95
Consensus MGFETGLEKCKADKTVPVELLSADGHSYHKSCHCKGTLVMSNYSSMDGVLYCKPHEQLFKETGSPKRNFPPTCAKA--NN--QSKVFNKRYGSV
```

X QT X E A V X I NK R Q I X Y YD KL P R X T RX AAM

### Domaines LIM2 et C-terminal des protéines des groupes $\alpha$ LIM1 et $\beta$ LIM1

```
TaWLim1 FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 196
HvWLim1 FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 196
OsWLim1 FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 195
ZmWLim1 FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 195
SoWLim1a FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 194
Consensus FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN DH DRTSQSA-----GSVDD DSEY
```

F N T D A A LDED EX

```
PtaGLIM1a FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 195
PtaGLIM1b FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 196
GhaLim1 FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 208
Ee $\alpha$ LIM1 FAGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DRTSQSA-----GSVDD--DSEY 194
Consensus FXGTREKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN ER ER KP TATEIAAES
```

AA I A R I KS S V T QP Q

```
NtWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-EMNST-----TTEVTAES-- 193
LeWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ETIPAI-----TTEVTAES-- 193
StLIM1b FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DKNSSV-----TTEVTAES-- 187
InWLim1a FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-EKNLV-----VMEIAAE-- 192
LsWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ERSA-----PG-QVAP-- 185
HaWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ERSA-----VG-ETAP-- 188
PtaWLim1a FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ERSA-----TTEVTAES-- 197
PtaWLim1b FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ERSA-----TTEVTAES-- 197
GhWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ETD-----GTEIAAES-- 190
EgLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-ERE-----TMAPES-- 188
EeWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-DSNTS-----EL-- 188
GmWLim1b FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-EKSATNGK-----INGEEVATET-- 196
AtWLim1 FGGTRKRCVGCSTVYPIERVVNNNTMYHKSCHCKGCTISPSNYIAHEGKLYCKHHHIQLIKERGNFSOLEN--DH-EKSATNGK-----KVVAA-- 190
```

# Annexes

BnWLM1 FGGTRKRCVGCCKTVYVPEKVVVNGTLYHKSCFKCTHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-DNAAKD-----KVDA----- 189  
 Consensus FXGTREKXCXCVTVYPTERVTVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG  
 D I T T K R K

LsPLIM1 FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----SAFAPAPES----- 189  
 ToPLIM1 FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----ST-SESTAESTPASAPASAPAS----- 201  
 HaPLIM1a FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----T-VAAPAESETQNTETQNAETQNAETQNAETQNTETQNSV 219  
 GhPLIM1a FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----T-VATP-ES----- 184  
 NtPLIM1a FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALST--QSLT----- 191  
 LpPLIM1 FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPP--QIAMES----- 193  
 PGPS/D1 FGGTRKRCVGCCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPTLQIAAES----- 195  
 Consensus FXGTREKXCXCVTVYPTERVTVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPP--QIAMES----- 195  
 TD XX LI K S K D P KX SR R -- -D

PtdβLIM1a FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-HHEVHT----PVKETAABE 191  
 PβLIM1b FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-EHLT----PVDETAAE 191  
 EtβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----LDN-HD-LG----- 179  
 GaβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENTVA- 189  
 GmβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENRITE 191  
 LeβLIM1a FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----HEKN- 179  
 LsβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----HEKIK- 189  
 NtβLIM1a FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----DRKIK- 189  
 InβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----EKSKE-V--PVTVNGTA- 191  
 LsβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----FMARKNFSQFDKQHEP-NHGVFNGVTENTTEV 196  
 TkβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----NGVTENTTEV 194  
 HaβLIM1 FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVNTENTTEV 188  
 Consensus FGGTQDKCAVCKKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVNTENTTEV 188  
 G E IX N I G XA RP K XY H X K M R X E --D

## Domaines LIM2 et C-terminal des protéines des groupes γLIM2 et δLIM2

TaWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQPTAQAAAQTPAPAAESS- 204  
 HwWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQPTAQAAAQTPAPAAESS- 204  
 SoWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----EQPSSDSS 197  
 ZmWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----EQPSSDSS 198  
 OsWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----APPPAADSS 196  
 Consensus FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP XQAAADSS  
 E S GQA T PP SE

StWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----PAATVPDS 189  
 LeWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----PAATVPDS 189  
 PhWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----PAATVPDS 189  
 NtWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----PAATVPDS 189  
 InWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----PTTVPDA 189  
 HpwLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----TAAATAAATAAASIEP 199  
 ToWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAASIAV 201  
 LsWLM2a FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 199  
 AtWLM2a FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 199  
 AtWLM2a FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 200  
 GmWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 198  
 PtaWLM2a FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 189  
 PtdWLM2b FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 189  
 EeWLM2 FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 189  
 GaWLM2b FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP-----SAAAATAAATAA 189  
 Consensus FSGTQDKCATCCKTAYPLEKVVVEKSYHKSFCFKSHGGCAITPSNYAALEGILYCKHHFSLFKERKGSYNHLIKCASVKRRAEAQP XQAAADSS  
 D X V I S I X V P A I DS

PtaPLIM2a FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-HHEVHT----PVKETAABE 184  
 PtaPLIM2b FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-EHLT----PVDETAAE 185  
 EePLIM2 FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----LDN-HD-LG----- 189  
 AtPLIM2b FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENTVA- 186  
 BnPLIM2 FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENRITE 186  
 AtPLIM2a FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALST--QSLT----- 191  
 GrPLIM2b FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPP--QIAMES----- 186  
 NtPLIM2 FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPTLQIAAES----- 187  
 HaPLIM2 FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPTLQIAAES----- 183  
 Consensus FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPTLQIAAES----- 183  
 C ST IQ SY S K X X S N N RV N L XKK

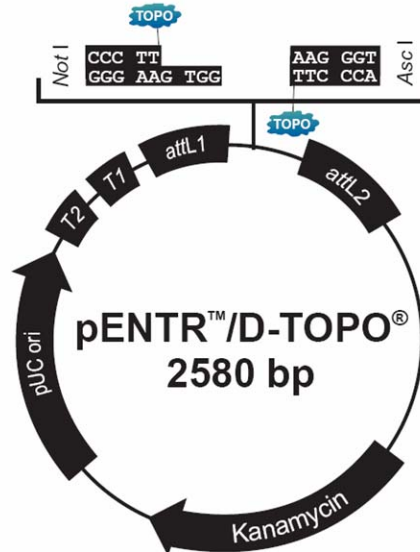
PtaPLIM2a FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----SAFAPAPES----- 180  
 OsPLIM2c FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----ST-SESTAESTPASAPASAPAS----- 182  
 TaPLIM2b FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----T-VAAPAESETQNTETQNAETQNAETQNAETQNTETQNSV 183  
 HvPLIM2 FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----T-VATP-ES----- 183  
 ZmPLIM2a FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----T-VAAPAESETQNTETQNAETQNAETQNAETQNTETQNSV 183  
 OsPLIM2b FGGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----T-VATP-ES----- 182  
 Consensus FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGGHDHETD PALPTLQIAAES----- 182  
 C L D T A A S I QN K NL - P A NXKE

PtδLIM2a FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-HHEVHT----PVKETAABE 216  
 PtδLIM2b FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-EHLT----PVDETAAE 212  
 ToδLIM2 FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----LDN-HD-LG----- 196  
 HaδLIM2b FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENTVA- 203  
 NtδLIM2 FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENRITE 201  
 LpδLIM2 FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----HEKN- 207  
 Consensus FSGTQDKCAACCKTVYVPEKVVVNGTAYHRAFCFKCHGGCTISPSNYIAHEGRLYCKHHHILQIKKGNLSQLEGG-E-----GVTENTTEV 207  
 E QA K X KI T N F A A Q NY -T -V X -XX

## Annexes

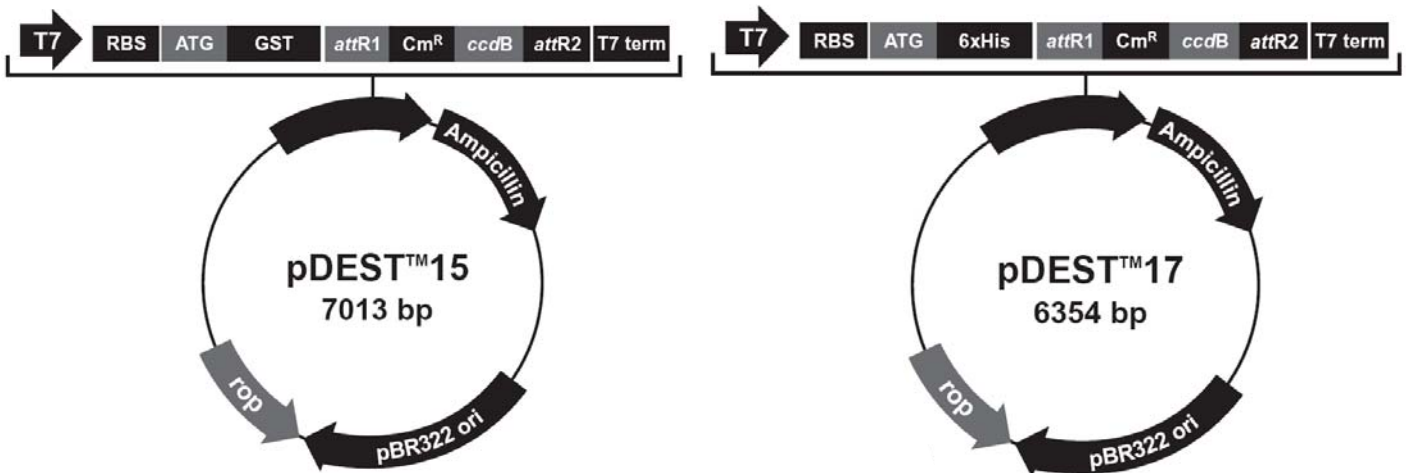
**Annexe 7.** Carte de la séquence du vecteur pENTR™/D-TOPO® utilisé pour le clonage de la séquence codante du gène *PtaGLIM1a*

Le plasmide comprend les séquences de terminaison de la transcription *rrnB* T1 et *rrnB* T2 (T1 et T2), les deux sites 1 (avec un bout collant) et 2 de reconnaissance (TOPO®) par la topoisomérase I entre les sites de recombinaison *attL1* et *attL2*, le gène de résistance à la kanamycine et l'origine de réplication pUC (pUC ori). Les sites des amorces sens et antisens M13, non représentés sur la figure, sont respectivement situés de part et d'autre des sites de recombinaison *attL1* et *attL2*.



**Annexe 8.** Carte de la séquence des vecteurs de destination pDEST™15 et pDEST™17 respectivement utilisé pour la production des protéines recombinantes GST-PtaGLIM1a et 6His-PtaGLIM1a

Le plasmide comprend le promoteur T7, le site de fixation par les ribosomes (RBS), le codon d'initiation de la traduction (ATG), les étiquettes 6-histidine (6xHis) pour le vecteur pDEST™17 ou glutathione S transférase (GST) pour le vecteur pDEST™15, puis entre les sites de recombinaison *attR1* et *attR2* le gène de résistance au chloramphénicol (*Cm<sup>R</sup>*) et le gène *ccdB* pour la sélection des bactéries DB3.1, la région de terminaison de la transcription T7 (T7 term), et enfin le gène *bla* de résistance à l'ampicilline, l'origine de réplication (pBR322 ori) et l'ORF *ROP* (*rop*) pour la réplication du vecteur à un faible nombre de copies.



## Annexes

## Genome-Wide Analysis of *LIM* Gene Family in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*

Dominique ARNAUD, Annabelle DÉJARDIN, Jean-Charles LEPLÉ, Marie-Claude LESAGE-DESCAUSES, and Gilles PILATE\*

*Equipe 'Xylème' Unité Amélioration, Génétique et Physiologie Forestières, INRA-Orléans, Avenue de la Pomme de Pin, BP 20619 Ardon, F-45166 Olivet Cedex, France*

(Received on 11 January 2007; accepted on May 12, 2007; published online 15 June 2007)

### Abstract

In Eukaryotes, LIM proteins act as developmental regulators in basic cellular processes such as regulating the transcription or organizing the cytoskeleton. The LIM domain protein family in plants has mainly been studied in sunflower and tobacco plants, where several of its members exhibit a specific pattern of expression in pollen. In this paper, we finely characterized in poplar six transcripts encoding these proteins. In *Populus trichocarpa* genome, the 12 *LIM* gene models identified all appear to be duplicated genes. In addition, we describe several new LIM domain proteins deduced from *Arabidopsis* and rice genomes, raising the number of *LIM* gene models to six for both species. Plant *LIM* genes have a core structure of four introns with highly conserved coding regions. We also identified new LIM domain proteins in several other species, and a phylogenetic analysis of plant LIM proteins reveals that they have undergone one or several duplication events during the evolution. We gathered several LIM protein members within new monophyletic groups. We propose to classify the plant LIM proteins into four groups:  $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2, and  $\delta$ LIM2, subdivided according to their specificity to a taxonomic class and/or to their tissue-specific expression. Our investigation of the structure of the LIM domain proteins revealed that they contain many conserved motifs potentially involved in their function.

**Key words:** poplar; *Arabidopsis*; rice; LIM domain protein; tension wood

### 1. Introduction

LIM proteins have been named by the initials of the three first discovered LIM homeodomain proteins: LIN11, ISL1, and MEC3.<sup>1–3</sup> In Eukaryotes, LIM proteins contain one or more LIM domains, which are, in some cases, associated with a protein kinase domain or a homeodomain. The LIM domain is a cysteine–histidine-rich, zinc-coordinating domain consisting of two zinc fingers repeated in tandem.<sup>4</sup> It is conserved over a wide variety of species. The cysteine-rich protein (CRP) family in animals is a subclass of LIM proteins characterized by their two LIM domains, with the consensus sequence [C-X<sub>2</sub>-C-X<sub>16–23</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>16–21</sub>-C-X<sub>2–3</sub>-(C/D/H)], that are both followed

by a short glycine-rich repeat. The CRP family in vertebrates involves four proteins: CRP1, CRP2, CRP3/MLP, and TLP, which act as molecular adapters.<sup>5</sup> Indeed, CRP1, CRP2, and CRP3 are able to bind  $\alpha$ -actin and zyxin, two components of the cytoskeleton. Although the LIM domains from animal CRP proteins are structurally similar to the GATA-type zinc finger transcription factor,<sup>6</sup> their DNA-binding activity is yet unproven and CRP may rather be involved in protein–protein interactions.

The LIM protein family in plants is CRP-related proteins containing two LIM domains separated by a long inter-LIM domain. On the contrary to the animal CRP family, plant LIM proteins have a longer C-terminal domain and lack the glycine-rich region (GRR) following each LIM domain. For all plant LIM proteins, the two LIM domains of 52 residues have the following characteristic structure: [C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-H].<sup>7</sup> The first gene encoding a LIM domain protein

Edited by Kazuo Shinozaki

\* To whom correspondence should be addressed. Tel. +33 2-38417875. Fax: +33 2-38417879. E-mail: pilate@orleans.inra.fr

© The Author 2007. Kazusa DNA Research Institute.

in plants has been named *SF3*. Later renamed *HaPLIM1*, *SF3* expression was found specific to sunflower pollen.<sup>8,9</sup>

The LIM proteins from sunflower, tobacco, and *Arabidopsis* have been classified into four groups: PLIM1 and PLIM2 specifically expressed in pollen and WLIM1 and WLIM2 expressed in the whole plant.<sup>7</sup> Like animal CRP, most plant LIM proteins are present in the cytoplasm and/or in the nucleus. This is the case for the sunflower protein HaWLIM1 that, for many different cell types, localizes either in the cytoplasm, in the nucleus, or in both.<sup>10</sup> Moreover, in protoplasts, HaWLIM1 seems to be associated with cortical microtubules, and it is also observed in the nucleus during the interphase.<sup>11</sup> As for CRP proteins, the tobacco NtWLIM1 binds F-actin and may be involved in actin cytoskeleton stability.<sup>12</sup> The sunflower protein HaPLIM1 has been detected both in small cytoplasmic structures located in the microspores and in the cortical region of mature pollen grains, where it concentrates in the actin-enriched germination cones, suggesting its interaction with the actin cytoskeleton.<sup>13</sup> Although HaPLIM1 was never found in the nucleus of vegetative cells, it exhibits a non-specific DNA- and RNA-binding activity.<sup>14</sup> Hence, the function of HaWLIM1 and HaPLIM1 in the transcriptional regulation remains unclear. In tobacco, NtLIM1 is clearly a transcription factor that binds to the PAL-box, a conserved motif present in the promoter of a number of genes from the phenylpropanoid pathway.<sup>15</sup> Transgenic tobacco plants with a reduced *NtLIM1* expression also present reduced lignin content and a decreased expression of PAL, 4CL, and CAD, three enzymes involved in the lignin biosynthesis. In poplar, the distribution of expressed sequence tags (ESTs) in different wood tissues indicates a rather high expression of an LIM protein homologue in tension wood.<sup>16</sup> Accordingly, microarray analyses also indicate a higher expression of some LIM transcription factor homologues in tension wood.<sup>17</sup> Tension wood, formed on the upper side of bent stems, is enriched in cellulose due to the formation of a supplementary gelatinous layer.

Our study focused on the plant LIM domain proteins containing only two LIM domains homologous to the animal CRP family. First, we finely describe, in this paper, the poplar *LIM* gene family. We determined the complete sequence of six cDNAs encoding LIM proteins and searched for LIM domain protein encoding genes in the *Populus trichocarpa* genome sequence. Secondly, we completed the inventory of the *Arabidopsis* and rice *LIM* gene family. To get a global overview of the plant LIM domain family, cDNAs and ESTs encoding LIM domain proteins have extensively been researched in plant sequence databases. Sequence analyses and phylogenetic studies revealed the structural diversity in plant LIM proteins. We named the genes coding for LIM domain proteins by following a nomenclature stemming from the phylogenetic analysis.

## 2. Materials and methods

### 2.1. Characterization of poplar LIM cDNA and gene sequences

The transcripts similar to LIM domain proteins were searched by basic local alignment search tool (BlastN) within a collection of 10 062 ESTs obtained from *Populus tremula* × *Populus alba* (clone INRA #717-1-B4, *Populus* section) wood cDNA libraries.<sup>16</sup> Among the *LIM* ESTs grouped in a same contig, we selected the longest cDNA clone and sequenced it. The entire cDNA was amplified by polymerase chain reaction (PCR) using the forward primer TriplexA and the reverse primer pTriplexB1<sup>16</sup> of the pTriplex vector (Clontech, Laboratories Inc., Mountain View, CA, USA). We further completed the forward and reverse sequencing with new primers specific to each *PtaLIM* gene (*Pta* stands for *P. tremula* × *P. alba*; Supplementary Table S1). The different *PtLIM* genes (*Pt* stands for *P. trichocarpa*, *Tacamahaca* section) were identified in the *Populus* genome database ([http://www.genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://www.genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html))<sup>18</sup> by keyword search, using the InterPro LIM domain annotation (IPR001781) as a query, and using the six cDNAs from *P. tremula* × *P. alba* as query for BlastN searches. We selected only the proteins containing two LIM domains. The prediction of exon/intron splicing in *P. trichocarpa* *LIM* genes was verified by sequencing reverse transcriptase (RT)-PCR fragments with specific primers against poplar *LIM* genes (see Supplementary Table S1). These fragments were produced after amplification of polyA RNA prepared from poplar wood samples.<sup>19</sup> Finally, the poplar ESTs homologous to the various *PtaLIM* transcripts identified were also searched against the GenBank database (<http://www.ncbi.nlm.nih.gov/>) and against the PopulusDB database (<http://www.populus.db.umu.se>), where cDNA libraries were built from *P. tremula* or *P. tremula* × *Populus tremuloides* (*Populus* section) samples.<sup>20</sup>

### 2.2. Database search for sequences coding for LIM domain protein

Until August 2006, we have used several ways to search in multiple databases all plant LIM proteins containing only two LIM domains. First, known genes and full-length cDNAs encoding LIM domain proteins from sunflower, tobacco, and *Arabidopsis* were collected from the literature and GenBank database.<sup>7</sup> Additional genes and full-length cDNA annotated as ‘LIM domain protein’, or ‘LIM transcription factor’ were found by keyword searches or using the InterPro LIM domain annotation (IPR001781) as a query. Alternatively, the poplar LIM protein sequences were used for BLAST (TblastN and BlastP) searches at the GenBank non-redundant database. The genomic and cDNA sequences of *Arabidopsis thaliana* and *Oryza sativa* were obtained

from the GenBank, TIGR plant genomic group (<http://plantgenomics.tigr.org/>), and TAIR (<http://arabidopsis.org>) databases. Finally, to get more sequences encoding LIM domain proteins, ESTs homologous to plant LIM domain proteins were searched by BlastN at the GenBank EST database. This process was repeated with each newly identified set of plant *LIM* genes until no further sequences with significant similarity were identified. For each gene, the longest EST was translated in all reading frames using the EMBOSS Transeq program at the EMBL–EBI, and only those carrying the entire coding sequence (CDS) and some part of 3' and 5' untranslated region (UTR) were chosen for further phylogenetic analyses. When EST sequences were too short or did not contain the entire CDS, a consensus sequence was deduced using the Bioedit software. Finally, the deduced amino acid sequences were verified for those carrying the entire two LIM domains using the PROSITE database (<http://www.expasy.org/prosite>). The selected ESTs with their consensus sequences are listed in Supplementary Table S2.

### 2.3. Sequence and phylogenetic analysis of LIM domain proteins

The selected protein sequences were aligned using the ClustalW software package (<http://www.ebi.ac.uk/clustalw/>)<sup>21</sup> with minor adjustments. Phylogenetic analyses were carried out using the Phylogenetic Interference Package (PHYLP) program, version 3.63 (<http://evolution.genetics.washington.edu/phylip.html>). Genetic distance matrices using protein polymorphism were calculated using PROTDIST software with the JTT amino acid substitution matrix as measure of distances.<sup>22</sup> A phylogenetic tree was then constructed using the neighbor-joining method thanks to NEIGHBOR software.<sup>23</sup> To estimate the statistical robustness of nodes, 1000 bootstrap samples were generated with SEQBOOT software, and the majority rule consensus tree was generated by CONSENSE software. The plant LIM family was also analyzed through a parsimonious method using the PROTPARS program with 1000 bootstrap replicates. Maximum likelihood analyses were performed using Phylml v2.4.1<sup>24</sup> with the JTT matrix and 100 bootstrap replicates. Maximum likelihood trees were generated with BIONJ, a modified neighbor-joining algorithm.<sup>25</sup> Trees were viewed and edited with Tree View,<sup>26</sup> and bootstrap values <50% were not reported.

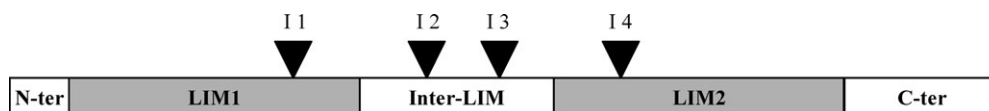
Conserved motifs in LIM domain proteins were detected using the ClustalW alignment with few manual corrections, and the MEME program (<http://www.meme.sdsc.edu/meme/meme.html>).<sup>27</sup> The aligned protein sequences were shaded using the Bioedit software with a threshold of 90% for identical residues and a BLOSUM62 matrix for shading similar residues. Isoelectric point (pI) and molecular weight (Mw) were predicted using the pI/Mw tool at expasy ([http://www.expasy.org/tools/pi\\_tool.html](http://www.expasy.org/tools/pi_tool.html)). PROSITE results were used to find putative ASN-glycosylation and phosphorylation sites.

## 3. Results and discussion

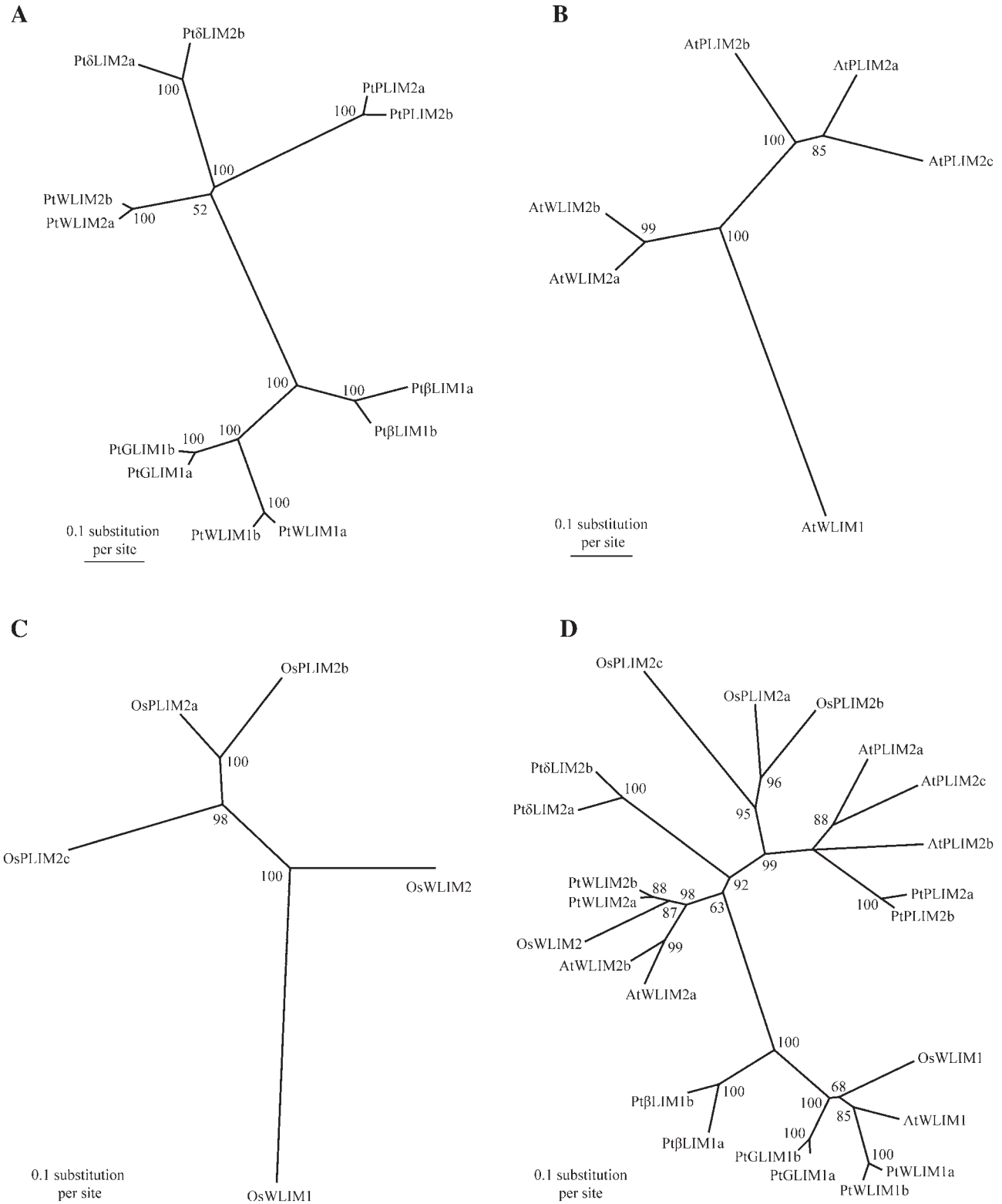
### 3.1. Survey and characterization of the LIM domain proteins in poplar, Arabidopsis and rice

**3.1.1. Isolation of the cDNAs coding for poplar LIM domain proteins and identification of 12 gene models in the *Populus trichocarpa* genome** The distribution of 10 062 poplar ESTs in different wood cDNA libraries<sup>16</sup> revealed that a cDNA, named *PtaGLIM1a* (EF035035), homologous to the LIM protein SF3<sup>8</sup> was abundant in differentiating xylem from tension wood. Five other *PtaLIM* cDNAs present in this EST collection were found, and a complete sequence was determined for each. We named them *PtaGLIM1b* (EF035036), *PtaWLIM2a* (EF035040), *PtaPLIM2a* (EF035037), *PtaWLIM1a* (EF035038), and *PtaWLIM1b* (EF035039). All of the full-length cDNAs obtained contain an entire CDS, a 5' and a 3' UTR, and encode protein with two LIM domains (Fig. 1; Supplementary Fig. S1).

We searched the *P. trichocarpa* genome sequence for all the gene models coding for LIM domain proteins. We excluded several gene models coding for proteins containing only one LIM domain linked to either a cytochrome P450 domain or an ubiquitin interaction motif and focused on gene models with two LIM domains. Besides the gene models corresponding to the six cDNAs isolated in our laboratory, we identified six other gene models encoding LIM domain proteins, raising to twelve the number of *LIM* gene models in the *Populus* genome. In accordance with their phylogenetic relationship with other known plant LIM domain proteins, we named these genes *PtWLIM1a*, *PtWLIM1b*, *PtGLIM1a*, *PtGLIM1b*, *PtBLIM1a*, *PtBLIM1b*, *PtWLIM2a*, *PtWLIM2b*, *PtPLIM2a*, *PtPLIM2b*, *PtδLIM2a*, and *PtδLIM2b* (Fig. 2A; Supplementary Table S3). With



**Figure 1.** Schematic diagram of the structure of the poplar LIM domain proteins. The conserved LIM1 and LIM2 domains are indicated by gray boxes. Open boxes show the N-terminal, inter-LIM, and C-terminal domains. Arrows indicate the respective position of the four introns (I1–I4).



**Figure 2.** Phylogenetic trees of *P. trichocarpa*, *A. thaliana*, and *O. sativa* LIM domain proteins. Four phylogenetic trees of (A) 12 poplar, (B) six *Arabidopsis*, (C) six rice, and (D) all deduced LIM domain proteins are shown. Amino acid sequences of LIM domain proteins were analyzed by neighbor-joining method with genetic distance calculated by the JTT model of amino acids change. The numbers at the nodes represent percent of bootstraps values ( $\geq 50\%$ ) based on 1000 replications. The length of the branches is proportional to the expected numbers of amino acid substitutions per site with a scale provided at the bottom of the trees. A species acronym is added before each LIM protein name: At, *A. thaliana*; Os, *O. sativa*; Pt, *P. trichocarpa*.



new information from both full-length *PtaLIM* cDNAs and *P. trichocarpa* transcript sequences,<sup>28</sup> we have corrected and improved the annotated poplar genomic sequence for the *PtLIM* genes (Supplementary Data 1).

Each pair of genes (*a* and *b*) exhibits a high sequence similarity from 85% amino acids identity between *Pt $\delta$ LIM2a* and *b* to 95% amino acids identity between *PtaGLIM1a* and *b* (Fig. 2A; Supplementary Table S4). This high similarity strongly suggests a gene duplication, as previously observed in a recent study for the poplar cellulose synthase *CesA* gene family.<sup>29</sup> These observations are in accordance with the hypothesis that poplar is an ancient polyploid and that a large-scale duplication event has occurred in the ancestor of poplar.<sup>18,30</sup> The duplicated LIM proteins may have kept similar functions, but the distribution of ESTs as well as the expression pattern between duplicated genes (data not shown) differs slightly, suggesting some other differences within their regulatory regions. All the deduced proteins have the features of the plant LIM domain protein family as described previously,<sup>7</sup> with two LIM domains separated by a long-spacer named the inter-LIM region, a short N-ter domain, and a C-ter domain variable in length (Supplementary Fig. S1). The length of *PtLIM* proteins is rather constant, between 194 and 216 amino acids, and their Mw varies between 20.9 and 24.1 kDa (data not shown). Poplar LIM proteins share a level of homology ranging from 43 to 95% of amino acid identity (Supplementary Table S4). The highest divergence between poplar LIM protein sequences is mainly localized to the inter-LIM and C-ter domains. From the single analysis of amino acid identity between the different poplar LIM domain proteins, we differentiated these proteins into two major groups: LIM1 and LIM2. Within each group, the percentage of identity at the amino acid level is a lot higher (from 58 to 95%) than between the two groups (from 43 to 55%).

**3.1.2. The Arabidopsis and rice genome contain six LIM gene models** A previous study reported the identification of three LIM genes in the *A. thaliana* genome: *AtWLIM1* (At1g10200), *AtPLIM2* (At2g45800), and *AtWLIM2* (At2g39900).<sup>7</sup> Because the sequences of both the *Arabidopsis* and rice genomes were publicly available, we had the opportunity to identify extensively all the genes coding for LIM domain proteins. Bioinformatics analyses performed against the GenBank and TIGR databases show that the *Arabidopsis* genome contains three other *AtLIM* genes. Because these genes seem to be duplicated, we named them *AtPLIM2b* (At1g01780), *AtPLIM2c* (At3g61230), and *AtWLIM2b* (At3g55770, also named *AtL2*<sup>31</sup>), whereas the previous *AtWLIM2* and *AtPLIM2* genes have been renamed *AtWLIM2a* and *AtPLIM2a*, respectively (Fig. 2B). The Supplementary Table S5 lists genomic, cDNA, and EST accessions for the six

*Arabidopsis LIM* genes, as well as the cDNA and genomic clones that contain errors or encode partial LIM domain proteins. It should be pointed out that the three related genes *AtPLIM2a*, *b*, and *c* overlap, respectively, with a gene encoding a phosphomannomutase (At2g45790), a gene encoding an unknown protein (At1g01770), and a gene encoding an oxydoreductase (At3g61220) that may affect the transcription of these *AtPLIM2* genes.

As in *Arabidopsis*, the rice (*Oryza sativa*) genome contains six genes coding for protein with two LIM domains: *OsWLIM1* (LOC\_Os12g32620), *OsWLIM2* (LOC\_Os03g15940), *OsPLIM2a* (LOC\_Os02g42820), *OsPLIM2b* (LOC\_Os04g45010), *OsPLIM2c* (LOC\_Os10g35930), and *OsLIM* (LOC\_Os06g13030) (Fig. 2C; Supplementary Table S5). *OsPLIM2a*, *b*, and *c* are very similar in their sequences and, therefore, may be considered in-paralogs genes. Unlike poplar, duplication is not the rule in *Arabidopsis* and rice (Fig. 2D). The genes *OsWLIM2*, *OsPLIM2b*, and *OsPLIM2c* are well supported by ESTs and are represented by full-length cDNAs.<sup>32</sup> For *OsWLIM1*, the identified cDNA (AK058220) is truncated. *OsPLIM2a* is also poorly represented at the mRNA level, with only two ESTs found and no published full-length cDNA. Therefore, for these two genes, we used their genomic sequences for the sequence alignment. In the case of *OsLIM*, we identified a very long transcript (AK102383) that encodes an unusual LIM protein of 1303 amino acids with two classical LIM domains and a very long C-ter domain that has no homology for any known protein. Only one (CI584223) of the 22 ESTs found by a BlastN search localizes in the 5' part of the transcript at the level of the first LIM domain. Because of its unusually long C-ter domain, we did not include this LIM sequence in the phylogenetic analysis.

### 3.1.3. Genomic organization of poplar, Arabidopsis, and rice genes

From the genomic analysis, we can infer that all plant LIM genes have a core gene structure with four introns within the coding sequence with the exception of *AtPLIM2a* and *c* genes that have two and three introns, respectively (Fig. 1; Supplementary Fig. S2). The position of the first and last introns is strictly conserved in the first and second LIM domain, respectively. In poplar and *Arabidopsis*, the *WLIM2* genes diverge from the other LIM genes by the occurrence of one (*AtWLIM2a*, *PtWLIM2a*, and *b*) or two (*AtWLIM2b*) supplementary introns in the 5' UTR, before the ATG initiation codon. For these *WLIM2* genes in eudicots species, a mechanism of alternative splicing of the first intron may be involved in post-translational regulation. This is supported by northern-blot experiments revealing two hybridizing bands for *AtWLIM2b* only in the shoot and not in the root.<sup>31</sup>

The length of coding regions CR1, CR2, and CR4 is highly conserved between plant LIM genes, indicative of

a strict conservation in the length of the LIM domains during evolution (Supplementary Fig. S2). The first coding region is 135–138 bp long and only *OsWLIM1*, *OsWLIM2*, and *OsLIM* have a variable CR1 length. The second coding region is 97 bp long for genes belonging to the LIM1 group and 100 bp long for genes from the LIM2 group. The fourth exon, localized within the second LIM domain, is the most conserved exon, with a length of 90 bp for all *LIM* genes. The CR3 and CR5 are highly variable in length resulting in the differences observed at the amino acid level, respectively, in the interLIM region and the C-ter domain. Finally, the *PLIM2* and  $\delta$ *LIM2* genes contain the longest fifth coding region that reflects the extensive length of the C-ter region of the deduced proteins.

**3.1.4. Identification of ESTs homologous to LIM domain protein** In a previous study, LIM proteins from sunflower, tobacco, and *Arabidopsis* have been classified into two groups, LIM1 and LIM2, and subdivided into four subgroups: PLIM1 and PLIM2, specifically expressed in pollen, and WLIM1 and WLIM2, widely expressed in plant.<sup>7</sup> Because of the availability of the genome sequence for *P. trichocarpa*, *A. thaliana*, and *O. sativa*, we found an increased number of genes belonging to the plant LIM domain family. The newly discovered LIM proteins may define new LIM subgroups or be related to the previously identified subgroups. To approach the diversity of the *LIM* gene family, an extensive BlastN search of cDNAs and ESTs encoding proteins with two LIM domains has been performed in NCBI plant sequence databases. In plants, we found 165 unigenes homologous to LIM domain protein, but we did not include 49 of them in the phylogenetic analysis because they contained only partial CDS (data not shown). Within the 116 unigenes coding for an entire LIM domain protein, 90 have a representative EST containing a full length CDS, whereas for each of the remaining 26 unigenes, a consensus was built to determine the complete CDS (Supplementary Table S2). We identified ESTs from a wide range of species in the different groups of

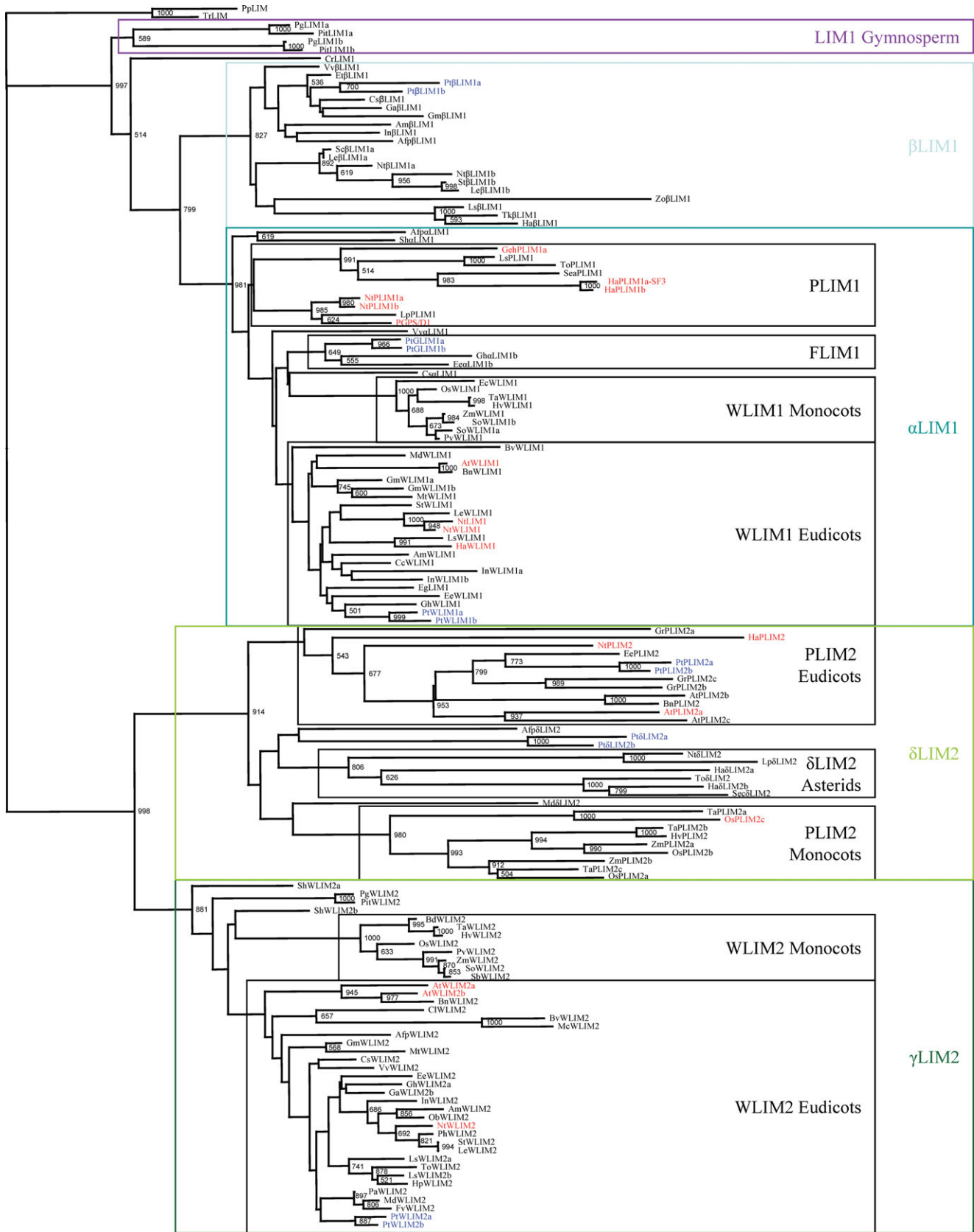
plants: bryophytes, conifers, piperales, monocotyledons, and eudicotyledons including rosids, asterids, Caryophyllales, and Ranunculales subclasses. In tobacco (*Nicotiana tabacum*) and in the sunflower (*Helianthus annuus*), the number of different *LIM* transcripts has been raised to eight with the finding of novel ESTs, notably the *HpWLIM2* EST from *Helianthus petiolaris*, whose existence was previously suspected in this genus by others.<sup>7</sup>

### 3.2. Phylogenetic analysis of plant LIM proteins with regard to expression data

**3.2.1. Four different groups** Phylogenetic trees have been constructed with the deduced amino acid sequence of the genes encoding LIM domain proteins (Fig. 3). We renamed the plant LIM proteins according to their phylogenetic relationship. The LIM1 and LIM2 groups identified previously<sup>7</sup> are clearly separated and supported by a high bootstrap value at the level of the TrLIM and PpLIM proteins from the mosses *Physcomitrella patens* and *Tortula ruralis* (Fig. 3). We were unable to place these two LIM proteins within either group because their sequence shares similarities with both LIM groups. Therefore, the phylogenetic trees have been rooted using sequences of these two Bryophyte LIM proteins. The plant LIM family can be divided into four groups,  $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2, and  $\delta$ LIM2 resulting from the division of the LIM1 and LIM2 groups. These four groups are supported by high bootstrap values. This phylogenetic analysis confirms the existence of the PLIM1, WLIM1, PLIM2, and WLIM2 subgroups as described previously.<sup>7</sup>  $\beta$ LIM1 is a new group that has not been identified before, whereas the  $\alpha$ LIM1 group includes PLIM1 and WLIM1 subgroups. The WLIM2 and PLIM2 subgroups belong, respectively, to the  $\gamma$ LIM2 and  $\delta$ LIM2 groups. In addition, each group contains new subgroups, which are described below.

**3.2.2. The  $\alpha$ LIM1 group** The previous WLIM1 and PLIM1 subgroups are gathered within the  $\alpha$ LIM1 group. With regards to the low bootstrap value

**Figure 3.** Phylogenetic tree of plant LIM domain proteins. Amino acid sequences of 149 LIM domain proteins were analyzed by neighbor-joining method with genetic distance calculated by the JTT model of amino acids change. The numbers at the nodes represent bootstraps values ( $\geq 500$ ) based on 1000 replications. The lengths of the branches are proportional to the expected numbers of amino acid substitutions per site with a scale provided at the bottom of the tree. LIM proteins have been renamed following their belonging to groups or subgroups WLIM1, PLIM1, FLIM1,  $\alpha$ LIM1,  $\beta$ LIM1, WLIM2, PLIM2, or  $\delta$ LIM2 that are in boxes. Poplar LIM proteins are in blue color, and plant LIM proteins whose expression pattern has been determined are in red color. A species acronym is added before each LIM protein name: Afp, *Aquilegia formosa*  $\times$  *A. pubescens*; Am, *Antirrhinum majus*; At, *A. thaliana*; Bd, *Brachypodium distachyon*; Bn, *Brassica napus*; Bv, *Beta vulgaris*; Cc, *Coffea canephora*; Cl, *Curcuma longa*; Cs, *Citrus sinensis*; Cr, *Ceratopteris richardii*; Ec, *Eleusine coracana*; Ee, *Euphorbia esula*; Et, *Euphorbia tirucalli*; Eg, *Eucalyptus globulus*; Fv, *Fragaria vesca*; Geh, *Gerbera hybrida*; Gm, *Glycine max*; Gh, *Gossypium hirsutum*; Gr, *Gossypium raimondii*; Ga, *Gossypium arboreum*; Ha, *Helianthus annuus*; Hp, *Helianthus petiolaris*; Hv, *Hordeum vulgare*; In, *Ipomoea nil*; Le, *Lycopersicon esculentum*; Lp, *Lycopersicon pennellii*; Ls, *Lactuca sativa*; Md, *Malus  $\times$  domestica*; Mt, *Medicago truncatula*; Mc, *Mesembryanthemum crystallinum*; Nt, *Nicotiana tabacum*; Ob, *Ocimum basilicum*; Pa, *Prunus armeniaca*; Pg, *Picea glauca*; Ph, *Petunia  $\times$  hybrida*; Pit, *Pinus taeda*; Pp, *Physcomitrella patens*; Pt, *Populus trichocarpa*; Pv, *Panicum virgatum*; Sea, *Senecio aethnensis*; Sec, *Senecio chrysanthemifolius*; Sb, *Sorghum bicolor*; Sh, *Saruma henryi*; So, *Saccharum officinarum*; St, *Solanum tuberosum*; Sc, *Solanum chacoense*; Ta, *Triticum aestivum*; To, *Taraxacum officinale*; Tk, *Taraxacum kok-saghyz*; Tr, *Tortula ruralis*; Vv, *Vitis vinifera*; Zo, *Zingiber officinale*; Zm, *Zea mays*. For additional details on each gene, see Supplementary Table S2.



0.1 substitutions per site

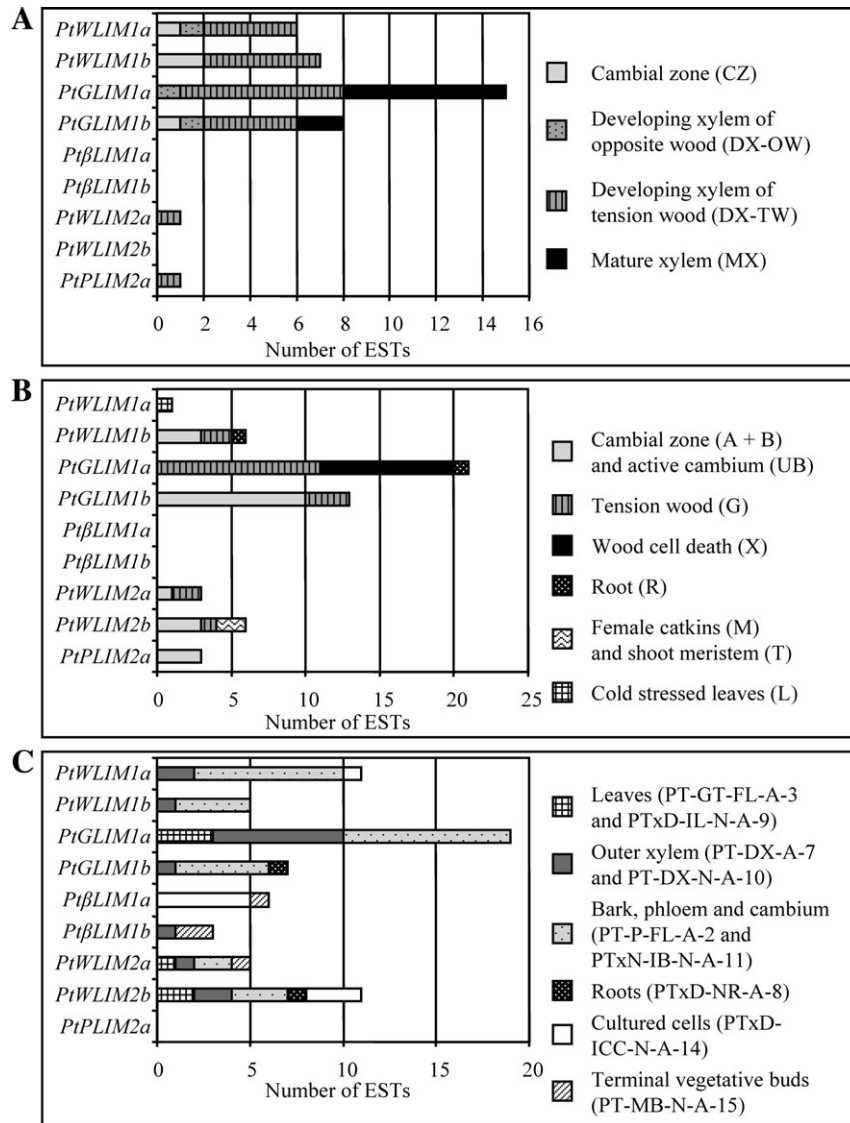
supporting the node, the definition of these two subgroups remains questionable (Fig. 3). In contrast, the monocots *WLIM1* subgroup clearly forms a statistically significant new monophyletic clade distinct from the eudicots *WLIM1* subgroup. Additionally, a fourth subgroup *FLIM1* could also be assigned to the  $\alpha$ *LIM1* group. However, the respective position of the *PLIM1*, *FLIM1*, and monocots and eudicots *WLIM1* subgroups within the  $\alpha$ *LIM1* group needs to be clarified. Indeed, the neighboring trees generated from a matrix of distances calculated by the JTT method or from the maximum likelihood (PhyML) method gave us an *FLIM1* subgroup close to the eudicots and monocots *WLIM1* subgroups and clearly separated from the *PLIM1* subgroup (Fig. 3; Supplementary Fig. S3). However, when using the parsimonious method, the *FLIM1* subgroup takes place within the *PLIM1* subgroup, between the *PLIM1* proteins from the Solanaceae and the Asteraceae families (Supplementary Fig. S3). Furthermore, this last method favors the hypothesis of a common ancestor for the monocots *WLIM1* and *PLIM1* subgroups. Presently, the *FLIM1* subgroup includes only a few proteins, namely *PtGLIM1a* and *b*, *Gh $\alpha$ LIM1*, and *Ee $\alpha$ LIM1*. Although additional sequences are definitely needed to strengthen this classification, there are strong arguments for the creation of a *FLIM1* subgroup: (i) *PtGLIM1a* and *b* share 81% of amino acid identity with *Gh $\alpha$ LIM1* and 79% with *Ee $\alpha$ LIM1*; (ii) RT-PCR analyses indicate that the *PtGLIM1a* and *b* genes are not expressed in pollen (data not shown), which could be an argument against grouping *PtGLIM1a* and *b* within the *PLIM1* subgroup. With 58 ESTs coming mostly from different xylem cDNA libraries (Fig. 4), *PtGLIM1a* is probably highly expressed in xylem tissues and to a lesser extent in other organs with vascular tissues, such as leaves. In contrast to *PtGLIM1a*, *PtGLIM1b* ESTs are preferentially found in cambial zone and phloem. The *PtGLIM1a* and *b* names include a G, indicative of their abundance in tension wood with G-fibers.<sup>16,17</sup> Similarly, all of the 31 *Gh $\alpha$ LIM1* ESTs were found in mature cotton fibers and a majority (29 ESTs) came from fibers harvested during secondary cell wall formation (Supplementary Table S2). Accordingly, we named this subgroup *FLIM1* with the F, indicative of the high expression of its members in fibers. We can speculate that during the evolution of poplar and cotton species, these three proteins may have gained a novel function important for fiber differentiation or maturation and more particularly in fibers containing a cellulose-enriched secondary cell wall.

Interestingly, the *PLIM1* subgroup contains only *LIM* proteins from Solanaceae and Asteraceae species from the asterids subclass (Fig. 3). Many expression studies have demonstrated the pollen-specific expression of these *PLIM1* genes. For example, the sunflower Ha*PLIM1a* (SF3) is a protein exclusively and highly expressed in maturing pollen,<sup>13</sup> and its colocalization with actin in

the germinating cone suggests that Ha*PLIM1a* participates in pollen germination or pollen tube elongation. The tobacco *NtPLIM1a* gene has also been shown to be highly expressed in mature pollen,<sup>33</sup> whereas in *Gerbera hybrida*, the expression of *GhPLIM1a* gene is highly upregulated in stamens as evidenced in microarray studies.<sup>34</sup> Likewise, the petunia *PGPS/D1* gene encoding a *PLIM1* protein has been found to be highly expressed in germinating pollen,<sup>35</sup> and the accumulation of the *PhPLIM1* transcript increases during anther maturation suggesting again an involvement of *PLIM1* proteins in pollen tube elongation. *PLIM1* proteins from Solanaceae species are clearly separated from those belonging to the Asteraceae species, suggesting that *PLIM1* proteins have diverged during the speciation of these two families.

The monocots *WLIM1* subgroup contains proteins with high homology (from 89 to 99% of amino acid identity). Monocots *WLIM1* ESTs originate from a large amount of tissue including inflorescences. Only the closely related *SoWLIM1a* and *b* genes from sugarcane seem to be duplicated, as the two maize *WLIM1* transcripts described are probably alleles of the same gene. Surprisingly, a high number of *WLIM1* ESTs has been sequenced from the maize (200) and sugarcane (128) cDNA libraries. In particular, *SoWLIM1a* and *b* with 76 ESTs is the biggest cluster of the 'shoot-root transition zone from adult plant' cDNA library.

All plant species from core eudicotyledons are represented within the large eudicots *WLIM1* subgroup. Within this subgroup, only poplar, *Ipomoea*, and soybean species have in-paralogs *WLIM1* proteins (Fig. 3). In general, each plant *WLIM1* gene is represented by a large number of ESTs originating from a wide variety of tissues, sometimes including inflorescence organs. The *WLIM1* genes have been the most studied *LIM* genes in plants, but their precise function is yet to be defined. In the sunflower, the Ha*WLIM1* protein is localized in the nucleus and/or the cytoplasm for different cell types,<sup>10</sup> and may be associated with the tubulin cytoskeleton in protoplasts.<sup>11</sup> A tobacco *WLIM1* gene, *NtLIM1*, highly expressed in the stem, acts as a transcription activator. Indeed, the binding of the *NtLIM1* protein to the PAL-box motif, present on the promoter of several genes from the lignin biosynthesis pathway, leads to the activation of the transcription of these genes. Moreover, an important reduction in the lignin content has been observed in antisense *Ntlim1* transgenic tobacco.<sup>15</sup> *NtLIM1* protein mostly differs from *NtWLIM1* protein<sup>7</sup> by seven additional residues in the C-ter of the protein. The *NtWLIM1* protein binds actin directly with a high affinity, enhances the stability of actin cytoskeleton, and promotes the bundling of actin filaments.<sup>12</sup> The poplar ESTs homologous to *PtWLIM1a* and *b* are more represented in vascular tissues with many ESTs from phloem, cambium, and xylem (Fig. 4). In *Eucalyptus globulus*, the homologous *EgLIM1* gene (AB208710) and



**Figure 4.** Distribution of poplar LIM EST in different poplar cDNA libraries. **(A)** Libraries from INRA-Orléans.<sup>16</sup> **(B)** Libraries from Umea Plant Science Centre (UPSC).<sup>20</sup> **(C)** Libraries from University of British Columbia (UBC).<sup>28</sup> cDNA libraries prepared from young and mature leaves, along with green shoot tips (PT-GT-FL-A-3), local mature leaves harvested after continuous feeding by *M. disstria* (PTxD-IL-N-A-9), outer xylem harvested biweekly between April and October 2002 (PT-DX-A-7 and PT-DX-N-A-10), phloem and cambium (PT-P-FL-A-2), bark (with phloem and cambium attached) harvested after continuous feeding by *C. lapathi* (PTxN-IB-N-A-11), roots harvested from 3-month-old trees grown in hydroponic media without nitrogen source for 24 and 48 h, as well as trees grown in regular media (PTxD-NR-A-8), cultured cells grown in media supplemented with salicylic acid, benzothiadazole, methyl jasmonate, chitosan, or *Pollacia radiosa* extract (PTxD-ICC-N-A-14); for more details, see Ralph et al.<sup>28</sup>

the corresponding cDNA (AB208709) have also been isolated, but there is no expression information available. We have been unable to position the Sh $\alpha$ LIM1, Afp $\alpha$ LIM1, Vva $\alpha$ LIM1, and Csa $\alpha$ LIM1 proteins in either subgroup of the  $\alpha$ LIM1 group, as illustrated by the low bootstrap values supporting the nodes.

In conifer trees, the LIM1 proteins are very similar in structure in white spruce (*Picea glauca*) and loblolly pine (*Pinus taeda*). For both species, the LIM1 proteins have undergone a probable duplication event, and it is difficult to class them into either the  $\alpha$ LIM1 or the  $\beta$ LIM1 subgroup (Fig. 3). In *Pinus taeda*, the *PitLIM1b* (clone

6C12H) gene is more highly expressed in the xylem than in any other part of the tree.<sup>36</sup> Therefore, at least in tree species, the LIM domain proteins from the WLIM1 subgroup, as well as those from the FLIM1 subgroup, are preferentially expressed during vascular development.

**3.2.3. The new  $\beta$ LIM1 group** The new  $\beta$ LIM1 group contains LIM proteins from species mostly belonging to the asterids and rosids subclasses, including the poplar duplicated Pt $\beta$ LIM1a and b proteins. Interestingly, there are no *Arabidopsis* or rice proteins into this group (Fig. 3). The proteins from Solanaceae and Asteraceae

form two distinct smaller clades within this group. Within the asterids subclass, only the Solanaceae species (tobacco, potato, and tomato) have in-paralogs  $\beta$ LIM1 proteins. The expression or function of the  $\beta$ LIM1 genes is yet unknown. Like the genes from the WLIM1 and WLIM2 subgroups,  $\beta$ LIM1 ESTs originate from a variety of tissues including flower and fruit (data not shown), but they are preferentially found in roots, in cells that undergo differentiation like immature cotton fibers, and in undifferentiated cells like callus. In poplar, the *Pt $\beta$ LIM1a* ESTs almost exclusively originate from cell suspension cultures and the *Pt $\beta$ LIM1b* ESTs are more widely represented in xylem, vegetative buds, and roots (Fig. 4). Although  $\beta$ LIM1 ESTs originate from all parts of the plant, a significant number has been sequenced from undifferentiated cell samples.

**3.2.4. The  $\gamma$ LIM2 group** The  $\gamma$ LIM2 group is clearly separated from the  $\delta$ LIM2 group with a high bootstrap value (Fig. 3). The  $\gamma$ LIM2 group contains the monocots and eudicots WLIM2 subgroups that are clearly separated. As in the WLIM1 subgroup, the short branches are indicative of a high conservation between WLIM2 protein sequences within each subgroup. In several eudicots such as *Arabidopsis*, poplar, cotton, and lettuce, *WLIM2* genes are duplicated, whereas this is never the case in monocot species. For both monocots and eudicots WLIM2 subgroups, a high number of ESTs from various tissues, including inflorescences, have been sequenced. This is the case in *Arabidopsis* for the *AtWLIM2a* gene<sup>7</sup> as well as in poplar, for the *PtWLIM2a* and *b* genes (Fig. 4). The wide range of expression and the high conservation of *WLIM2* genes suggest that they may be involved in basic cellular processes. This is substantiated by the difficulty to produce homozygous mutants for *AtWLIM2a* gene.<sup>37</sup> In a recent microarray study, *AtWLIM2a* appeared highly expressed in siliques, and its expression was induced in the *Arabidopsis pkl* mutant.<sup>37</sup> The PKL protein is a negative regulator of transcription that acts as a repressor of embryonic identity during seed germination, and *AtWLIM2a* gene may be involved in the regulation of embryo development. In another study, *AtWLIM2a* expression is also gradually induced during leaf senescence.<sup>38</sup> Taken together, the *AtWLIM2a* gene may be more precisely involved in the seed maturation. The duplicate of *AtWLIM2a*, *AtWLIM2b* (*AtL2*), is more highly expressed in roots than in shoots. In addition, this expression seems to be affected by nitrogen availability.<sup>31</sup>

**3.2.5. The  $\delta$ LIM2 group** Because of the increased number of LIM sequences available, the  $\delta$ LIM2 group is considerably enlarged. The  $\delta$ LIM2 group is divided into three monophyletic subgroups: the eudicots PLIM2, monocots PLIM2, and Asterids  $\delta$ LIM2 that contains the sequences from the new LIM domain proteins identified

in sunflower and tobacco plants (Fig. 3). Several duplication events have occurred within the  $\delta$ LIM2 genes, indicative of an important diversification within this group. Because of the long length of branches, the  $\delta$ LIM2 proteins are on average more divergent each other than proteins from any other groups. Most eudicots *PLIM2* genes appear preferentially expressed during pollen development. This is the case for *AtPLIM2a* expressed in hydrated pollen grains,<sup>39</sup> for the tobacco *NtPLIM2*, and the sunflower *HaPLIM2* genes.<sup>7</sup> Likewise, in *Gerbera hybrida*, the ESTs encoding an homologue to *HaPLIM2* also appeared exclusively expressed in stamens<sup>34</sup> and in tomato, all the EST sequences orthologous to the *NtPLIM2* gene originated from a pollen cDNA library (data not shown). Surprisingly, a poplar cDNA library prepared from male catkins<sup>20</sup> does not contain any *PtPLIM2a* and *b* EST, but we observed for the *PtPLIM2a* gene a strong expression in mature anthers (data not shown). It should be noted that a few *PtPLIM2a* ESTs originate from cambial zone, tension wood, and root poplar libraries (Fig. 4), but no *PtPLIM2b* EST has been sequenced yet.

Monocots PLIM2 proteins are phylogenetically separated from eudicots PLIM2 proteins supported by a high bootstrap value. In monocots, the PLIM2 proteins are generally found in triplicate, resulting from at least two duplication events. Contrary to the highly conserved monocots WLIM2 proteins, monocots PLIM2 proteins are more divergent. As for eudicots, the monocots *PLIM2* genes are generally highly expressed in pollen, but not exclusively. Of the 200 *ZmPLIM2a* ESTs found in public databases, 156 come from flower cDNA libraries, and 66 from these ESTs originate from mature pollen. The majority of *OsPLIM2c* and *OsPLIM2b* ESTs also come from flower cDNA libraries. *OsPLIM2c* (AK072520) expression has been studied using microarray analysis during pollination and early embryogenesis in the rice pistil.<sup>40</sup> Their maximal accumulation in the pistil occurs at anthesis and decreases gradually during the following 24 h, when the pollen tube reaches the micropyle. However, the same study reveals a wider expression of the *OsPLIM2c* gene, suggesting an involvement in some other processes than pollen tube elongation. We speculate that within monocots and rosids species, the extensive duplication of their *PLIM2* genes may be a counterpart to the lack of specialization of the genes from the  $\alpha$ LIM1 group toward pollen-specific expression.

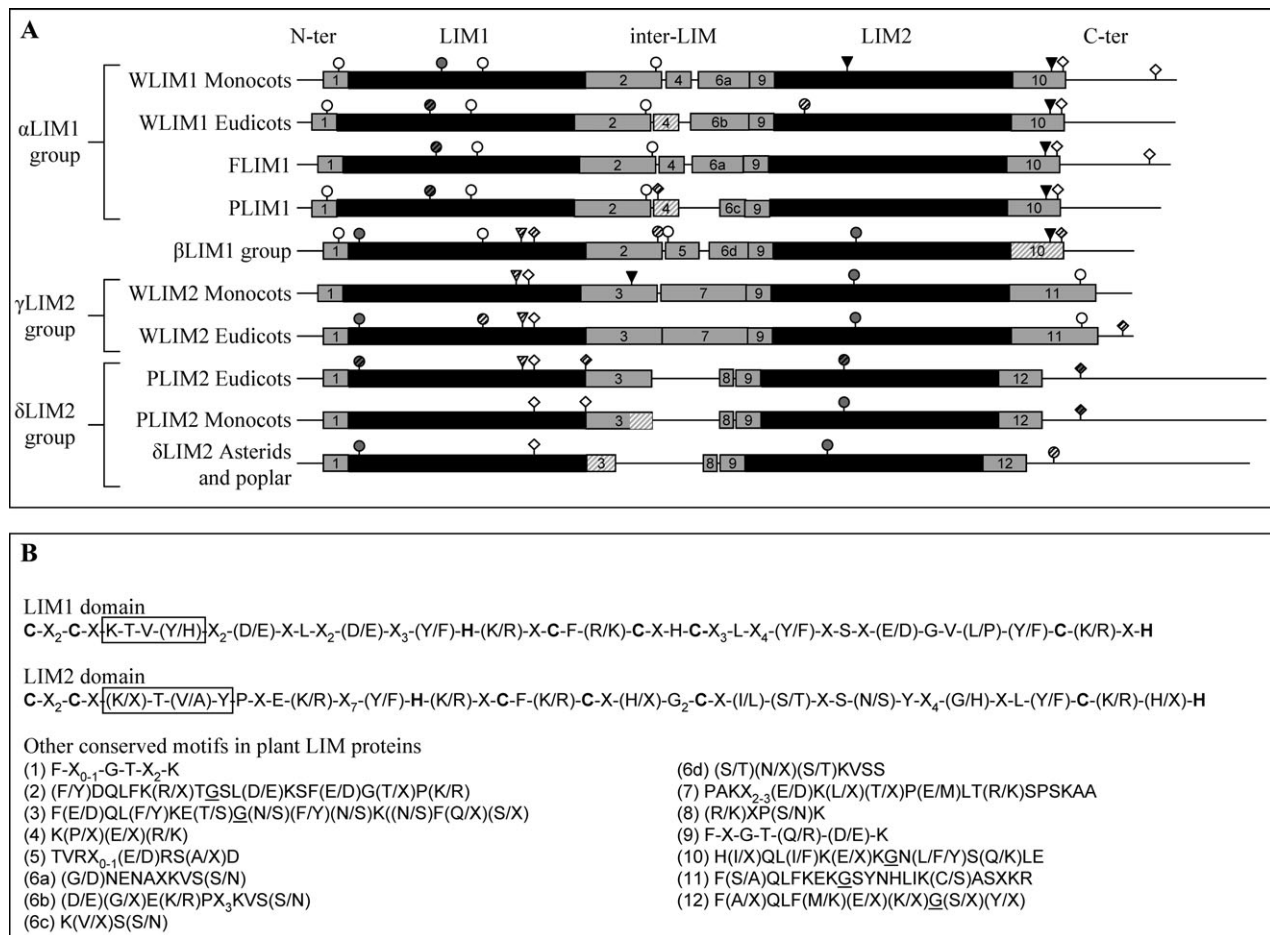
In asterids species, *LIM* genes also seem duplicated within the  $\delta$ LIM2 group. Contrary to the *Arabidopsis PLIM2* genes found in triplicate within the PLIM2 subgroup, the newly discovered tobacco and sunflower  $\delta$ LIM2 proteins are more distantly related from the previously identified HaPLIM2 and NtPLIM2 and therefore belong to a new Asterids  $\delta$ LIM2 subgroup (Fig. 3). The Asterids  $\delta$ LIM2 subgroup is surrounded by the two monocots and eudicots PLIM2 subgroups. Because dicots and

monocots *PLIM2* genes are strongly expressed in pollen, it is probable that these asterids  $\delta$ *LIM2* genes are also strongly expressed in pollen, but it is not clear if these duplication events have brought any new functions in pollen development or in other processes. Poplar *Pt* $\delta$ *LIM2a* and *b* protein sequences deduced from the *P. trichocarpa* genome are phylogenetically distant from the *PLIM2* subgroups. So far there is no expression data for their potential pollen-specific expression.

### 3.3. Sequence and structural analysis of plant LIM domain protein

**3.3.1. Toward a new characterization of the two LIM domains consensus sequences in plant** The plant LIM domain proteins present a lot of common features with the CRP proteins, but also several specificities. This is the case for the GRR following each LIM domain found in the animal CRP proteins that lacks in all the plant LIM domain proteins.<sup>7</sup> However, for all plant LIM domain proteins, a glycine residue is strictly conserved

nine amino acids after the last zinc ligand for each LIM domain (Fig. 5B). The potential nuclear targeting signal (KKYGPK) present in the CRP family of animal LIM proteins is also missing in the plant LIM domain family. The LIM domains contain two zinc finger repeated in tandem with the characteristic structure [C-X<sub>2</sub>-C-X<sub>17</sub>-H-X<sub>2</sub>-C]-X<sub>2</sub>-[C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-H] (Fig. 5).<sup>7</sup> Compared with the CRP family in animals, the second LIM domain in plants is atypical with the second ligand of the second zinc finger replaced by a conserved glycine. It has been proposed that closer histidine or cysteine makes the formation of two alternative structures -C-X-H-X<sub>18</sub>-C-X<sub>2</sub>-H- or -C-X<sub>4</sub>-C-X<sub>15</sub>-C-X<sub>2</sub>-H- possible for the zinc coordination.<sup>10</sup> However, the increased number of protein sequences analyzed reveals that this histidine residue is sometimes lacking in the LIM proteins from the  $\beta$ *LIM1* and  $\delta$ *LIM2* groups (Supplementary Fig. S4). Only the cysteine residue is conserved in all plant LIM domain proteins, strongly suggesting that the -C-X<sub>4</sub>-C-X<sub>15</sub>-C-X<sub>2</sub>-H- is the most probable structure for the



**Figure 5.** (A) Diagrammatic representation of motifs found in the sequences of the plant LIM family proteins. The LIM1 and LIM2 domains are represented by black boxes. ASN glycosylation sites are shown in triangles. cAMP- and cGMP-dependent protein kinase (cAMP), CKII, Tyr, and PKC phosphorylation sites are, respectively, shown in grey and white diamonds, grey and white circles. Non-conserved motifs are striped. Conserved motifs surrounding the LIM domains and identified with ClustalW<sup>21</sup> and MEME<sup>27</sup> programs are in gray boxes. (B) Consensus sequences of the two plant LIM domains and the 15 motifs identified. Cysteine and histidine coordinating zinc finger are highlighted in bold. The last conserved glycine after each LIM domain is underlined and the conserved K[T/A]VY motif is boxed. X represents any amino acid residue.

second zinc finger. Interestingly, the first cysteine of the first LIM domain is missing only for the sugarcane-duplicated SoWLIM1b protein and in maize for the second allelic version of ZmWLIM1 protein supported by the cDNA AY112454 (data not shown). If this cysteine is compulsory for the formation of the zinc finger, the functional significance of this absence remains to be elucidated. In both animal CRP and plant LIM families, with the exception of the monocots PLIM2 protein, a highly conserved K[T/A]VY motif is found in the first LIM domain close to the second cysteine of the first zinc finger (Fig. 5B). This common motif may be important in the correct folding of the LIM1 domain, and is also found, but conserved to a lesser extent, at the same position in the second LIM domain. In general, the amino acids surrounding the LIM domains are also highly conserved between LIM proteins of each subgroup with the noticeable exception of the  $\delta$ LIM2 proteins (Fig. 5A and B). It should be pointed out that the LIM1 domain is generally more conserved than the LIM2 domain. However, in WLIM2 proteins, both LIM domains are highly conserved.

*3.3.2. Structural variation may reflect functional differences* As for animals, plant LIM proteins may have a function in the regulation of cell differentiation or development. LIM proteins could have multiple partners in the cell, such as actin or tubulin cytoskeleton, DNA and other potential proteins. It has been proposed that plant LIM proteins may serve as a shuttle between the cytoplasm and the nucleus for yet unknown functions. In this respect, we searched for putative phosphorylation and glycosylation sites in the LIM protein sequences using the Prosite database. Thanks to the high number of sequences collected, we have been able to perform ortholog sequence comparisons within subgroups to find conserved motifs using the MEME program. Most of the differences observed between these LIM proteins are concentrated either in the inter-LIM region or in the C-ter domain (Fig. 5). Within the  $\alpha$ LIM1 group, PLIM1 proteins are more divergent each other than WLIM1 and FLIM1 proteins. The C-ter domain of PLIM1 proteins is more different than in WLIM1 proteins and contains additional proline residues (Supplementary Fig. S4). FLIM1 proteins mostly resemble WLIM1 proteins, particularly at the level of the LIM domains. Their inter-LIM domain is more similar to those of monocots WLIM1 proteins, whereas their C-ter domain differs clearly from that of the other subgroups. Proteins within the  $\beta$ LIM1 group carry on their LIM domains, the structural properties characteristic of both WLIM2 and  $\alpha$ LIM1 groups, but the inter-LIM and C-ter domains are specific to this group. WLIM2 proteins have highly conserved sequences, even at the level of the inter-LIM region, except for WLIM2 proteins from conifers and Caryophyllales. The  $\delta$ LIM2 group is the most heterogeneous

LIM group even at the level of LIM domains. The  $\delta$ LIM2 proteins are characterized by a long C-ter domain highly variable in amino acid composition, and composed of many acidic amino acids, particularly glutamic acid (Supplementary Fig. S4). It has been suggested for tobacco NtLIM1 protein that this acidic domain may function as a transcription activator.<sup>15</sup> Interestingly, on the contrary to PLIM2 and  $\alpha$ LIM1 proteins, the C-ter domain of WLIM2 proteins is rather basic and contains non-polar amino acids except one conserved acidic amino acid at the end of the protein. The C-ter domain of  $\beta$ LIM1 proteins has a variable *pI*. The *pI* is rather basic in the Solanaceae-duplicated proteins, and acid to neutral for the Asteraceae proteins (data not shown). Different kinds of motifs like the ASN glycosylation site, the casein kinase II (CKII), the tyrosine (Tyr), and the protein kinase C (PKC) phosphorylation sites are localized inside and outside the LIM domains (Fig. 5). For example, in  $\alpha$ LIM1 and  $\beta$ LIM1 proteins and only in these proteins, a putative PKC phosphorylation site is found just before the first zinc ligand of the LIM1 domain. However, the FLIM1 proteins (from the  $\alpha$ LIM1 group) do not possess this phosphorylation site. Finally, another PKC phosphorylation site is exclusively found at the C-ter domain in WLIM2 proteins from angiosperm and also gymnosperm species, suggesting a key function for this site.

#### 3.4. Concluding remarks

In this work, we report on the first analysis of poplar LIM domain protein family with the characterization of six *PtaLIM* cDNAs. Moreover, we updated the *LIM* genomic sequences in *P. trichocarpa*, *A. thaliana*, and *O. sativa*. The set of ESTs collected and the subsequent phylogenetic classification of LIM proteins provide a useful database for future researches on this family. We have defined four phylogenetic groups,  $\alpha$ LIM1,  $\beta$ LIM1,  $\gamma$ LIM2, and  $\delta$ LIM2, and positioned into these groups, the previously characterized PLIM1, WLIM1, PLIM2 and WLIM2 subgroups. The  $\beta$ LIM1 is a newly identified group, whose genes have an apparent preferential expression in undifferentiated cells. However, more detailed expression studies are needed for the functional analysis of these  $\beta$ LIM1 proteins. Besides, the pollen-specific expression of *PLIM1* and *PLIM2* genes, several subgroups seem to be specific to a plant class or subclass. This is the case for the PLIM1 subgroup that is represented only by LIM protein from the asterids subclass. Additionally, it appears that LIM proteins from monocots and eudicots plant are phylogenetically distant, suggesting specific functions within each taxonomic class. Finally, a new FLIM1 subgroup was created within the  $\alpha$ LIM1 group, and the corresponding genes appear to be highly expressed in poplar G-fibers or cotton fibers. In poplar, the distribution of ESTs



corresponding to *WLIM1* and *GLIM1* genes suggests the involvement of LIM proteins in wood fibers formation and/or vascular development. Despite the fact that plant LIM proteins could both bind to actin and act as a transcription factor, the function of LIM proteins remains largely unknown. Structural analysis is the first step to the functional characterization of LIM proteins in plants. In the future, we will need to determine whether all plant LIM proteins are able to bind actin cytoskeleton and/or DNA, and to examine the function of PLIM1 and PLIM2 proteins in pollen development or pollen tube elongation.

**Acknowledgements:** We are grateful to J.-F. Arnaud and V. Castric (Laboratoire de Génétique et Evolution des Populations Végétales, Université de Lille 1, France) for their advice in phylogenetic analysis, and I. Bourgait for her help in bioinformatics analysis. We also thank Kory Wein, Assistant Professor at the University of Platteville (WI, USA) for the English editing. D. Arnaud was supported by a fellowship grant from the Conseil Régional de la Région Centre. **Supplementary Data:** Supplementary data are available online at <http://dnaresearch.oxfordjournals.org>.

## References

1. Freyd, G., Kim, S. K. and Horvitz, H. R. 1990, Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-II*, *Nature*, **344**, 876–879.
2. Karlsson, O., Thor, S., Norberg, T., Ohlsson, H. and Edlund, T. 1990, Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys–His domain, *Nature*, **344**, 879–882.
3. Way, J. C. and Chalfie, M. 1988, *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*, *Cell*, **54**, 5–16.
4. Dawid, I. B., Breen, J. J. and Toyama, R. 1998, LIM domains: multiple roles as adapters and functional modifiers in protein interactions, *Trends Genet.*, **14**, 156–162.
5. Weiskirchen, R. and Günther, K. 2003, The CRP/MLP/TLP family of LIM domain proteins: acting by connecting, *BioEssays*, **25**, 152–162.
6. Perez-Alvarado, G. C., Miles, C., Michelsen, J. W., et al. 1994, Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP, *Nat. Struct. Biol.*, **1**, 388–398.
7. Eliasson, Å., Gass, N., Mundel, C., et al. 2000, Molecular and expression analysis of a LIM protein gene family from flowering plants, *Mol. Gen. Genet.*, **264**, 257–267.
8. Baltz, R., Domon, C., Pillay, D. and Steinmetz, A. 1992, Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein, *Plant J.*, **2**, 713–721.
9. Baltz, R., Evrard, J. L., Domon, C. and Steinmetz, A. 1992, A LIM motif is present in a pollen-specific protein, *Plant Cell*, **4**, 1465–1466.
10. Mundel, C., Baltz, R., Eliasson, A., et al. 2000, A LIM-domain protein from sunflower is localized to the cytoplasm and/or nucleus in a wide variety of tissues and is associated with the phragmoplast in dividing cells, *Plant Mol. Biol.*, **42**, 291–302.
11. Briere, C., Bordel, A.-C., Barthou, H., et al. 2003, Is the LIM-domain protein HaWLIM1 associated with cortical microtubules in sunflower protoplasts?, *Plant Cell Physiol.*, **44**, 1055–1063.
12. Thomas, C., Hoffmann, C., Dieterle, M., Van Troys, M., Ampe, C. and Steinmetz, A. 2006, Tobacco WLIM1 is a novel F-Actin binding protein involved in actin cytoskeleton remodeling, *Plant Cell*, **18**, 2194–2206.
13. Baltz, R., Schmit, A.-C., Kohnen, M., Hentges, F. and Steinmetz, A. 1999, Differential localization of the LIM domain protein PLIM-1 in microspores and mature pollen grains from sunflower, *Sex. Plant Reprod.*, **12**, 60–65.
14. Baltz, R., Evrard, J.-L., Bourdon, V. and Steinmetz, A. 1996, The pollen-specific LIM protein PLIM-1 from sunflower binds nucleic acids in vitro, *Sex. Plant Reprod.*, **9**, 264–268.
15. Kawaoka, A., Kaothien, P., Yoshida, K., Endo, S., Yamada, K. and Ebinuma, H. 2000, Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis, *Plant J.*, **22**, 289–301.
16. Déjardin, A., Leplé, J.-C., Lesage-Descauses, M.-C., Costa, G. and Pilate, G. 2004, Expressed sequence tags from poplar wood tissues—a comparative analysis from multiple libraries, *Plant Biol.*, **6**, 55–64.
17. Andersson-Gunneras, S., Mellerowicz, E. J., Love, J., et al. 2006, Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis, *Plant J.*, **45**, 144–165.
18. Tuskan, G. A., DiFazio, S., Jansson, S., et al. 2006, The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray), *Science*, **313**, 1596–1604.
19. Lafarguette, F., Leplé, J.-C., Déjardin, A., et al. 2004, Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood, *New Phytol.*, **164**, 107–121.
20. Sterky, F., Bhalerao, R. R., Unneberg, P., et al. 2004, A *Populus* EST resource for plant functional genomics, *Proc. Natl. Acad. Sci. USA*, **101**, 13951–13956.
21. Chenna, R., Sugawara, H., Koike, T., et al. 2003, Multiple sequence alignment with the Clustal series of programs, *Nucleic Acids Res.*, **31**, 3497–3500.
22. Jones, D. T., Taylor, W. R. and Thornton, J. M. 1992, The rapid generation of mutation data matrices from protein sequences, *Comput. Appl. Biosci.*, **8**, 275–282.
23. Saitou, N. and Nei, M. 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.*, **4**, 406–425.
24. Guindon, S. and Gascuel, O. 2003, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst. Biol.*, **52**, 696–704.
25. Gascuel, O. 1997, BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data, *Mol. Biol. Evol.*, **14**, 685–695.
26. Page, R. D. M. 1996, Tree View: an application to display phylogenetic trees on personal computers, *Comput. Appl. Biosci.*, **12**, 357–358.
27. Bailey, T. L. and Elkan, C. 1994, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.

28. Ralph, S., Oddy, C., Cooper, D., et al. 2006, Genomics of hybrid poplar (*Populus trichocarpa* × *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar, *Mol. Ecol.*, **15**, 1275–1297.
29. Djerbi, S., Lindskog, M., Arvestad, L., Sterky, F. and Teeri, T. 2005, The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (*CesA*) genes, *Planta*, **221**, 739–746.
30. Sterck, L., Rombauts, S., Jansson, S., Sterky, F., Rouze, P. and Van de Peer, Y. 2005, EST data suggest that poplar is an ancient polyploid, *New Phytol.*, **167**, 165–170.
31. Tranbarger, T. J., Al-Ghazi, Y., Muller, B., Teyssendier de la Serve, B., Doumas, P. and Touraine, B. 2003, Transcription factor genes with expression correlated to nitrate-related root plasticity of *Arabidopsis thaliana*, *Plant Cell Environ.*, **26**, 459–469.
32. Kikuchi, S., Satoh, K., Nagata, T., et al. 2003, Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice, *Science*, **301**, 376–379.
33. Sweetman, J., Spurr, C., Eliasson, A., Gass, N., Steinmetz, A. and Twell, D. 2000, Isolation and characterisation of two pollen-specific LIM domain protein cDNAs from *Nicotiana tabacum*, *Sex. Plant Reprod.*, **12**, 339–345.
34. Laitinen, R. A. E., Immanen, J., Auvinen, P., et al. 2005, Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae), *Genome Res.*, **15**, 475–486.
35. Guyon, V. N., Astwood, J. D., Garner, E. C., Dunker, A. K. and Taylor, L. P. 2000, Isolation and characterization of cDNAs expressed in the early stages of flavonol-induced pollen germination in *Petunia*, *Plant Physiol.*, **123**, 699–710.
36. Allona, I., Quinn, M., Shoop, E., et al. 1998, Analysis of xylem formation in pine by cDNA sequencing, *Proc. Natl. Acad. Sci. USA*, **95**, 9693–9698.
37. Dean Rider, S., Henderson, J. T., Jerome, R. E., Edenberg, H. J., Romero-Severson, J. and Ogas, J. 2003, Coordinate repression of regulators of embryonic identity by *PICKLE* during germination in *Arabidopsis*, *Plant J.*, **35**, 33–43.
38. Gepstein, S., Sabehi, G., Carp, M.-J., et al. 2003, Large-scale identification of leaf senescence-associated genes, *Plant J.*, **36**, 629–642.
39. Becker, J. D., Boavida, L. C., Carneiro, J., Haury, M. and Feijo, J. A. 2003, Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome, *Plant Physiol.*, **133**, 713–725.
40. Yoshida, K., Endo, M., Nakazono, M., et al. 2005, cDNA microarray analysis of gene expression changes during pollination, pollen-tube elongation, fertilization, and early embryogenesis in rice pistils, *Sex. Plant Reprod.*, **17**, 269–275.

## **Identification et caractérisation de la famille de protéines à domaine LIM chez le peuplier**

Les gènes *LIM* de plantes sont exprimés dans tous les tissus ou spécifiquement dans le pollen. Les protéines à domaine LIM de plantes auraient une double fonction cellulaire agissant comme facteur de transcription activant la synthèse de la lignine ou comme protéine liant et empaquetant les filaments d'actines. Ce travail de thèse présente la première étude de la famille de protéines à domaine LIM chez le peuplier. Six paires de gènes *LIM* ont été identifiées dans le génome séquencé de *Populus trichocarpa*, et la comparaison avec les modèles de gènes *LIM* d'*Arabidopsis* et de riz révèle une structure génomique très conservée. L'analyse de 155 unigènes *LIM* additionnels chez les plantes a permis de définir quatre groupes phylogénétiques distincts divisés en plusieurs sous-groupes monophylétiques différant par des spécificités d'expression tissulaires et/ou par des spécificités taxonomiques. L'analyse de l'expression de ces gènes par RT-PCR semi quantitative dans différents tissus et organes de peuplier montre que certains gènes sont préférentiellement exprimés dans les tissus vasculaires, alors que d'autres sont fortement exprimés dans le pollen ou dans les fibres de coton. La protéine PtaGLIM1a qui définit un nouveau sous-groupe FLIM1, est fortement exprimée dans le bois de tension. De plus, cette protéine subirait des modifications post-traductionnelles particulières dans le bois. Enfin, ces résultats suggèrent un rôle des protéines LIM de plantes dans certains aspects de l'expansion cellulaire.

**Mots clés :** Peuplier, *Arabidopsis*, Riz, Protéines LIM, Phylogénie, Analyse d'expression, Bois de tension

## **Identification and characterization of the poplar LIM domain protein family**

Plant *LIM* genes are expressed either in the whole plant or specifically in pollen. Plant LIM domain proteins may have a dual function as transcription factor activating lignin biosynthesis and/or as an actin binding and bundling protein. This PhD thesis work presents the first study on the LIM domain protein family in poplar. Six pairs of *LIM* genes have been identified in the sequenced *Populus trichocarpa* genome, and comparison with the gene models from *Arabidopsis* and rice show a highly conserved genomic structure. The analysis of 155 additional *LIM* unigenes from plants allows us to define four distinct phylogenetic groups divided into several monophyletic subgroups that differ by their tissue expression specificities and/or their taxonomic specificities. The semi quantitative RT-PCR expression analysis of these genes in different poplar tissues reveals that several genes are preferentially expressed in vascular tissues whereas others are highly expressed in pollen or in cotton fibers. The PtaGLIM1a protein that defines a new FLIM1 subgroup is highly expressed in tension wood. Moreover, this protein could undergo specific post-translational modifications in wood. Finally, these results suggest that plant LIM domain proteins may control some aspects of cell expansion.

**Keywords:** Poplar, *Arabidopsis*, Rice, LIM domain protein, Phylogeny, Expression analysis  
Tension wood

**Discipline :** Physiologie Génétique Moléculaire

### **Intitulé et adresse du laboratoire :**

INRA-Orléans, Unité Amélioration, Génétique et Physiologie Forestières  
Avenue de la pomme de pin, BP 20619 Ardon, 45166 OLIVET cedex, FRANCE