



**HAL**  
open science

# Contrôle transcriptionnel du développement de la graine des légumineuses. Caractérisation fonctionnelle du gène **LEC1-LIKE**

Jérôme Verdier

► **To cite this version:**

Jérôme Verdier. Contrôle transcriptionnel du développement de la graine des légumineuses. Caractérisation fonctionnelle du gène LEC1-LIKE. Sciences du Vivant [q-bio]. Université de Bourgogne, 2008. Français. NNT: . tel-02816951

**HAL Id: tel-02816951**

**<https://hal.inrae.fr/tel-02816951v1>**

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.

# THESE

Présentée pour l'obtention du

DIPLOME DE DOCTORAT

par

Jérôme VERDIER

**Contrôle transcriptionnel du développement de la  
graine des légumineuses.**

**Caractérisation fonctionnelle du gène *LEC1-LIKE***

Spécialité : Biochimie et Biologie moléculaire

Soutenue publiquement le 24 septembre 2008 devant le jury composé de :

Daniel WIPF

Martine DEVIC

Gilles CHARMET

Julia BUITINK

Bertrand DUBREUCQ

Richard THOMPSON

Professeur

Directeur de recherche CNRS

Directeur de recherche INRA

Chargé de recherche INRA

Chargé de recherche INRA

Directeur de recherche INRA

Président

Rapporteur

Rapporteur

Examineur

Examineur

Directeur de thèse



# REMERCIEMENTS

Bien qu'ayant principalement travaillé dans le domaine de la régulation génétique, je n'en reste pas moins conscient de l'importance de l'environnement sur le développement d'une plante ou d'un individu. Je tiens donc à remercier toutes les personnes travaillant à l'UMR Génétique et Ecophysiologie des Légumineuses qui m'ont permis d'évoluer dans un environnement scientifique et humain propice au bon déroulement de ce travail de thèse.

Je tiens à remercier tout particulièrement mon directeur de thèse Richard Thompson pour m'avoir offert l'opportunité de travailler dans son unité. Sa rigueur scientifique, son expertise, sa disponibilité et sa patience ont fait de cette thèse une expérience très enrichissante. J'ai apprécié la richesse de vos idées et la confiance que vous m'avez accordées.

Je remercie également Karine Gallardo et Christine Le Signor, mes co-encadrantes, qui m'ont guidé, conseillé et soutenu avec le sourire et beaucoup d'enthousiasme durant toutes ces années. Vos conseils et vos encouragements m'ont été particulièrement précieux.

Mes remerciements s'adressent également à tous les membres du jury : Gilles Charmet et Martine Devic, les rapporteurs de ce travail de thèse, qui ont gentiment accepté de se plonger dans ce travail pendant leurs vacances (désolé !); Daniel Wipf qui s'est volontiers acquitté de la fonction de Président ; Julia Buitink et Bertrand Dubreucq qui ont également accepté d'examiner ce travail. Je suis très flatté de votre présence dans ce jury et de votre disponibilité.

Je n'oublie pas les personnes qui ont participé au comité de suivi de cette thèse : Dominique Job, Loïc Lepiniec, Gilles Charmet, Peter Rogowski et Francis Marty. Merci pour tous ces conseils. J'ai beaucoup apprécié ces réunions annuelles faites dans la convivialité mais qui étaient scientifiquement très enrichissantes.

Un grand merci aux personnes qui m'ont aidé à réaliser ce travail : Grégoire Aubert pour tes conseils (presque toujours) avisés en biologie moléculaire mais surtout pour ton soutien ; Myriam Sanchez pour ton aide dans le suivi de mutation (presque toujours fiable !) et surtout pour ta bonne humeur ; Françoise Moussy et Brigitte Darchy pour votre aide en serre ; Delphine Hélicher pour les nombreux "coups de main" que tu m'as donné en protéomique ; Judith Burstin pour tes idées et conseils toujours judicieux ; Sergio Ochatt et Louis Jacas pour votre expertise en régénération *in vitro* des plantes ; Florence Cassecuelle pour ton aide en FPLC ; l'équipe du CRB Gérard Duc, Jean Bernard Magin-Robert et Marianne Martinello ; Anne-Lise Santoni pour les analyses Dumas ; Valérie Labas pour les identifications de spots protéiques et surtout Samia Djennane qui a initié ce travail.

Bien sur, je n'oublie pas mes deux compères du SOBBQ, Michaël et Vincent. C'est une belle page qui se tourne pour chacun d'entre nous mais je retiendrai nos discussions enflammées (rapport au BBQ), et surtout notre amitié qui a rendu cette expérience humainement inoubliable. Dernier message, au sujet de notre resto à fromage dans le Jura, je propose qu'on se laisse un an et on en reparle !

Je remercie également toutes les personnes du laboratoire qui ont contribué à la bonne humeur générale : Hélène, ma collègue de paillasse ; Jean, Françoise et Anthony pour ces moments de convivialités au café ; Damien pour m'avoir enseigné le "presque" papillon ; Amandine et Christelle du bureau du bout du monde ; Mona ; les secrétaires (les deux Dominique(s), Corinne et Francine) toujours partantes pour les coups "fumeux" ; et toutes les personnes qui sont passées dans l'unité : Tola, Pedro, Marie, Aurélie, Gaëlle, Sissi, ....

Enfin, mon dernier remerciement s'adresse à mes parents et à toi, Vanessa qui m'a portée et surtout supportée dans cette étape, comme durant toutes ces années. C'est ta présence et ton soutien au quotidien qui ont fait que ce travail a pu aboutir, merci pour tout ce que tu m'apportes.





## **Publications :**

- **Verdier J**, Kakar K, Gallardo K, Le Signor C, Aubert G, Schlereth A, Town CD, Udvardi MK, Thompson RD. (2008) Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol Biol.* 67(6):567-580.
- Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, **Verdier J**, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK. (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* 2008 Jun 4.
- **Verdier J** and Thompson RD (2008) Overview of transcriptional regulation during seed maturation in dicotyledons. *Plant and Cell Physiology*, review accepted.
- **Verdier J**, Le Signor C, Moussy F, Sanchez M, Hélicher D, Bendahmane A, Karine Gallardo and Richard Thompson. *LEAFY COTYLEDON1-LIKE*, a regulator of seed storage protein synthesis in pea (in prep).

## **Communications orales :** (les résumés des communications sont disponibles en **Annexe 1**)

- Etude des facteurs de transcription lors du développement de la graine de *M. truncatula* : cinétiques d'expression et compartimentation tissulaire. Graines 2007 (Angers) 7-8 juin 2007.
- Transcriptome de la morphogenèse et du remplissage de la graine. Groupe Protéagineux (Dijon) 14-15 juin 2007.
- Etude des facteurs de transcription lors du développement de la graine de *M. truncatula*. XIIIème forum des jeunes chercheurs (Dijon) 14-15 juin 2007.
- Gene expression dynamics during *M. truncatula* seed formation. International Congress of Genetics and Genomics (Brisbane, Australia) April 9-13, 2006.
- Real-time RT-PCR profiling of over 1400 *Medicago truncatula* transcription factors during seed filling. 3rd International Congress of Legume Genetics and Genomics, Brisbane, Australia (9-13 avril 2006).
- Régulation transcriptionnelle dans les graines des légumineuses. Journée Jeunes Chercheurs du DGAP, Dijon, France (20-21 octobre 2005).



### **Posters :**

- Real-time RT-PCR profiling of over 1400 *Medicago truncatula* transcription factors during seed filling. XII ème forum des Jeunes Chercheurs, Besançon, France (8-9 juin 2006).
- Real-time RT-PCR profiling of all annotated *Medicago truncatula* transcription factors during seed filling. International Model Legume Congress, Asilomar, California (5-9 juin 2005).

### **Communications orales présentant une partie de ce travail de thèse :** (Résumé en annexe 1)

- Use of *Medicago truncatula* to investigate interactions between maternal and filial tissues during legume seed development. Karine Gallardo, Jérôme Verdier, Christian Firnhaber, Hélène Zuber, Klementina Kakar, Delphine Hericher, Maya Belghazi, Helge Küster, Michael Udvardi and **Richard Thompson**. Model Legume Congress 2007.
- Genome-wide analysis of transcription factor gene expression by high-throughput qRT-PCR in *Medicago*. **Kakar, Klementina** , Wandrey, Maren, Cheung, Foo, Xiao, Yongli, Wu, Hank, Town, Christopher, Schlereth, Armin, Czechowski, Tomasz, Stitt, Mark , Verdier, Jerome, Thompson, Richard, Torres-Jerez, Ivone, Udvardi, Michael K. Plant Biology Congress 2007, Chicago, USA.



# TABLE DES MATIERES

## INTRODUCTION : Contexte, objectifs et démarche expérimentale

1. PRÉSENTATION DU SECTEUR PROTÉAGINEUX.....	2
1.1 Présentation du secteur	2
1.1.1 Production	2
1.1.2 Consommation : l'Europe, vers une autonomie en protéines ?	3
1.2 L'amélioration du secteur protéagineux : l'origine et les lignes directrices	4
2. PLANTES D'ÉTUDE.....	5
2.1 Le pois : plante d'intérêt agronomique	5
2.2 <i>M. truncatula</i> : plante modèle des légumineuses	6
3. LE DÉVELOPPEMENT DE LA GRAINE DES LÉGUMINEUSES.....	7
3.1 Formation des trois tissus de la graine	7
3.2 Synthèse et accumulation des protéines de réserve	9
4. CONTRÔLE TRANSCRIPTIONNEL DE LA SYNTHÈSE DES PROTÉINES DE RÉSERVE (Revue : Transcriptional regulation of Storage proteins during dicotyledon seed filling).....	10
5. OBJECTIFS ET DÉMARCHE DE LA THÈSE.....	30
5.1 Objectifs de la thèse	30
5.2 Démarche de ce travail de thèse	30

## CHAPITRE 1 : Exploration des gènes de *M. truncatula* exprimés lors du développement de la graine

1. OBJECTIFS ET DÉMARCHE.....	33
2. UN ATLAS D'EXPRESSION DE GÈNES CHEZ LA LÉGUMINEUSE MODÈLE <i>M. TRUNCATULA</i> (article : A gene expression atlas of the model legume <i>Medicago truncatula</i> ).....	35
3. ANALYSE DU TRANSCRIPTOME LORS DU DÉVELOPPEMENT DE LA GRAINE.....	46
3.1 Stratégie utilisée	46
3.2 Analyse des profils d'expression pendant le développement de la graine	46
3.3 Classification fonctionnelle des gènes différemment exprimés	47
3.4 Processus associés aux différentes étapes du développement	48
3.4.1 L'embryogénèse	48
3.4.2 Le remplissage du grain	49
3.4.3 La dessiccation	50
4. CONCLUSION DU CHAPITRE.....	52



## **CHAPITRE 2 : Identification de facteurs de transcription impliqués dans le remplissage de la graine de *M. truncatula* et isolement de mutants pour ces gènes**

1. OBJECTIFS ET DÉMARCHE.....	54
2. LE CONTRÔLE TRANSCRIPTIONNEL DU DÉVELOPPEMENT DE LA GRAINE DE <i>M. TRUNCATULA</i> (Article : Gene expression profiling of <i>M. truncatula</i> transcription factors identifies putative regulators of grain legume seed filling).....	56
3. SELECTION DE FACTEURS DE TRANSCRIPTION CANDIDATS IMPLIQUÉS DANS LE DÉVELOPPEMENT DE LA GRAINE DES LÉGUMINEUSES.....	71
3.1 Comparaison des données d'expression obtenues en PCR quantitative et celles obtenues par puce à ADN	71
3.2 Sélection de candidats expressionnels pour l'approche de validation	71
3.2.1 Critères de sélection	71
3.2.2 Candidats sélectionnés chez <i>M. truncatula</i>	72
3.3 Transfert des données chez le pois	73
4. ISOLEMENT DE MUTANTS POUR LES GÈNES CANDIDATS SÉLECTIONNÉS.....	74
4.1 Outils de génomique fonctionnelle disponibles chez <i>M. truncatula</i>	74
4.1.1 La mutagenèse	74
4.1.1.1 Population EMS/TILLING	74
4.1.1.2 Population Fast Neutron et rayons $\gamma$	74
4.1.2 La transgénèse	75
4.1.2.1 RNA interférence (PTGS)	75
4.1.2.2 Mutagenèse insertionnelle	75
4.1.2.3 Utilisation d'éléments transposables	76
4.2 Mutants obtenus	76
5. CONCLUSION DU CHAPITRE.....	78
Synthèse des résultats obtenus	
Perspectives de ce travail	

## **CHAPITRE 3 : Les gènes de type LEC1 chez *M. truncatula***

1. OBJECTIFS ET DÉMARCHE.....	81
2. INTRODUCTION.....	82
2.1 Choix des gènes de type LEC1 comme candidats	82
2.2 Caractéristiques des gènes <i>LEAFY COTYLEDON</i>	82
3. RÉSULTATS.....	83
3.1 Identification de deux gènes de type LEC1 chez <i>M. truncatula</i>	83
3.1.1 Comparaison de séquence	84
3.1.2 Analyse des profils d'expression	85
3.2 Recherche de mutants "perte de fonction" pour ces gènes chez <i>M. truncatula</i>	85
3.2.1 Criblage TILLING	85
3.2.2 Transformants RNAi	86
3.3 Analyse fonctionnelle du gène Mtl1L	87
3.3.1 Analyse des plantes transformées	88
3.3.1.1 Expression du gène	





3.3.1.2 Phénotypage	
3.3.1.3 Analyses biochimiques	
3.3.2 Analyse des mutants EMS/TILLING	89
3.3.2.1 Phénotypage des plantes mutées	
3.3.2.2 Analyses biochimiques	
4. DISCUSSION.....	90
4.1 Deux gènes de type LEC1 identifiés chez <i>M. truncatula</i>	90
4.2 Comparaison des résultats obtenus sur l'analyse des mutants	91
4.3 Rôle de MtL1L dans le développement de la graine	93
4.4 Hypothèses sur le rôle des gènes de type LEC1	94
5. CONCLUSION DU CHAPITRE.....	96
Synthèse des résultats obtenus	
Perspectives de ce travail	

## **CHAPITRE 4 : Rôle du gène PsL1L dans l'élaboration de la qualité de la graine de pois (*P. sativum*)**

1. OBJECTIFS ET DÉMARCHE.....	97
2. IDENTIFICATION ET CARACTÉRISATION FONCTIONNELLE DU GÈNE LEAFY COTYLEDON1-LIKE CHEZ LE POIS (article : LEAFY COTYLEDON1-LIKE, a regulator of seed storage protein synthesis in pea).....	99
3. ETUDE DU POLYMORPHISME DE SÉQUENCE DE PsL1L DANS UNE COLLECTION DE 372 ÉCOTYPES CONTRASTÉS DE POIS.....	117
3.1 Choix de la méthode d'EcoTILLING et présentation de la collection CRB de pois	117
3.1.1 Méthode d'EcoTILLING	117
3.1.2 Collection CRB d'écotypes de pois	118
3.2 Analyse du polymorphisme de séquence de PsL1L	118
3.3 Génétique d'association entre le SNP #4 et des caractères de qualité de la graine	119
3.3.1 Génétique d'association	119
3.3.2 Localisation géographique des différents allèles du SNP #4	120
3.3.3 Phénotypage et structuration des plantes portant l'allèle "minoritaire" du SNP #4	120
3.4 Implication de l'allèle lié au SNP #4 dans l'amélioration d'un caractère agronomique de pois	121
4. CONCLUSION DU CHAPITRE.....	122

## **CONCLUSION GÉNÉRALE 125**



## **MATÉRIELS ET MÉTHODES**

1. MATÉRIEL BIOLOGIQUE.....	129
1.1 Matériel végétal et conditions de culture	129
1.2 Souches bactériennes et milieux de culture	129
1.3 Plasmides utilisés	130
2. MÉTHODES D'ANALYSES DES ACIDES NUCLÉIQUES.....	131
2.1 Extraction des acides nucléiques	131
2.1.1 Extraction d'ADN	
2.1.2 Extraction d'ADN plasmidique	
2.1.3 Extraction d'ARN totaux	
2.2 Electrophorèse des acides nucléiques	133
2.2.1 Electrophorèse d'ADN	
2.2.2 Electrophorèse d'ARN	
2.3 Clonage moléculaire	134
2.4 Transcription inverse	134
2.5 Amplification génique (PCR)	135
2.6 RT-PCR quantitative	136
2.7 Hybridation In Situ	137
2.8 Séquençage des acides nucléiques	138
2.9 RACE-PCR	138
3. MÉTHODES D'ANALYSES DES PROTÉINES.....	139
3.1 Extraction des protéines totales	139
3.2 Dosage des protéines	139
3.3 Electrophorèse monodimensionnelle (1D)	139
3.4 Electrophorèse bidimensionnelle (2D)	140
3.4.1 Principe	
3.4.2 Isoélectrofocalisation	
3.4.3 Equilibration des strips	
3.4.4 Séparation sur gels SDS PAGE et coloration	
3.4.5 Analyse des gels 2D	
3.4.6 Identification des spots protéiques	
3.5 FPLC	144
4. MÉTHODE DE TRANSGENÈSE.....	144
4.1 Construction des vecteurs de transformation	144
4.2 Transformation et sélection des plantes	144
5. MÉTHODE TILLING ET ECOTILLING.....	145
5.1 TILLING	145
5.2 EcoTILLING	146
6. ANALYSES BIOINFORMATIQUES.....	146

## **RÉFÉRENCES BIBLIOGRAPHIQUES** **147**

### **ANNEXES**

Annexe 1 (Résumés de conférences).....	165
Annexe 2 .....	171
Annexe 3 .....	172



## Abbréviations

A, C, G, T	adénosine, cytosine, guanine, thymidine
aa	acide aminé
ABA	acide abscissique
ADN	acide désoxyribonucléique
ADNc	acide désoxyribonucléique complémentaire
ADN-T	acide désoxyribonucléique transféré
ARN	acide ribonucléique
ARNm	acide ribonucléique messenger
ARNr	acide ribonucléique ribosomal
BAC	"bacterial artificial chromosome" (chromosome artificiel de bactérie)
BeT	bromure d'éthidium
BSA	"bovine serum albumine"
CHAPS	3- [3-cholamidopr opyl-diméthylammonio]- 1 -pr opane- sul fonate
CTAB	cétyltriméthylammonium
CO <sub>2</sub>	dioxyde de carbone
DA, kDA	dalton, kilodalton
DEPC	diéthyl-pyrocarbonate
DIG	digoxygénine
dNTP	déoxynucléotides triphosphates densité optique
DO	densité optique
E, µE	einstein, microeinstein
EDTA	éthylène diamine tétraacétique
EMS	éthyl-méthylsulfonate
EST	"expressed sequence tag"
g, mg, µg	gramme, milligramme, microgramme
GA	acide gibbérélique
GUS	13-glucuronidase
h, min, s	heure, minute, seconde
IPTG	isopropyl-f3-D-thiogalactoside
JAP	jour après pollinisation
l, ml, µl	litre, millilitre, microlitre
LB	Luria Broth
LC-MS/MS	liquid chromatography mass spectrometry
M, mM, nM	molaire, millimolaire, micromolaire
MALDI-TOF	matrix-assisted laser-desorption ionization time of flight
MOPS	acide 3-(N-Morpholino)propane sulfonique
MS	Murashige et Skoog
pb, kb	paire de bases, kilobase
PBS	"phosphate-buffered saline"
PCR	polymerase chain reaction
pH	potentiel hydrogène
qPCR (qRT-PCR)	PCR quantitative ("quantitative reverse transcription PCR")
RE	réticulum endoplasmique
RNase	ribonucléase
RT	"reverse transcription" (rétrotranscription ou transcription inverse)
SDS	sodium dodécylsulfate (lauryl sulfate)
SSP	seed storage protein
TAE	tris acétate EDTA
TBS	"tris-buffered saline"
TEA	triéthanolamine
TF	facteur de transcription
U	unité enzymatique commerciale
UV	ultraviolets
V, kV	volt, kilovolt



# **INTRODUCTION : Contexte, objectifs et démarche expérimentale**





# INTRODUCTION

## Contexte, objectifs et démarche expérimentale du travail de thèse

Les graines ont toujours fourni une grande part de l'alimentation des peuples et sont même à l'origine de l'apparition des grandes cultures. Aujourd'hui, les graines permettent d'assurer à elles seules 70 % de l'alimentation humaine. Cette estimation se révèle encore plus importante lorsque l'on considère qu'elles servent également à l'alimentation animale. Jusqu'à récemment, les graines étaient simplement prélevées sur la récolte précédente, il n'existait pas de marché des semences. L'émergence d'une véritable filière « semences » a permis de sécuriser les récoltes et d'améliorer les rendements. Cela a également entraîné l'apparition d'un marché national qui, avec le progrès des transports, est devenu mondial. De part l'importance des graines dans l'alimentation, le marché mondial des semences a doublé depuis ces vingt dernières années et va certainement continuer à se développer en adéquation avec la démographie planétaire qui devrait atteindre plus de 9 milliards de personnes d'ici 2050, soit plus de 30 % d'augmentation par rapport à aujourd'hui.

Selon les dernières estimations de la Fédération Internationale des Semences (ISF), la valeur totale des semences utilisées dans le monde peut être évaluée à 34 milliards de dollars américains. En termes de marché commercial intérieur, sept pays dépassent le milliard de dollars : les États-Unis (7), la Chine (4), la France (1,9), le Japon, la Russie, le Brésil et l'Allemagne. Prise dans son ensemble, l'Union Européenne talonne toutefois les États-Unis, avec un marché intérieur de 7 milliards de dollars. Ces marchés sont relativement stables dans les pays développés. En revanche, ils connaissent une croissance importante dans les pays émergents (Chine, Brésil, Inde, Argentine, Mexique,...).

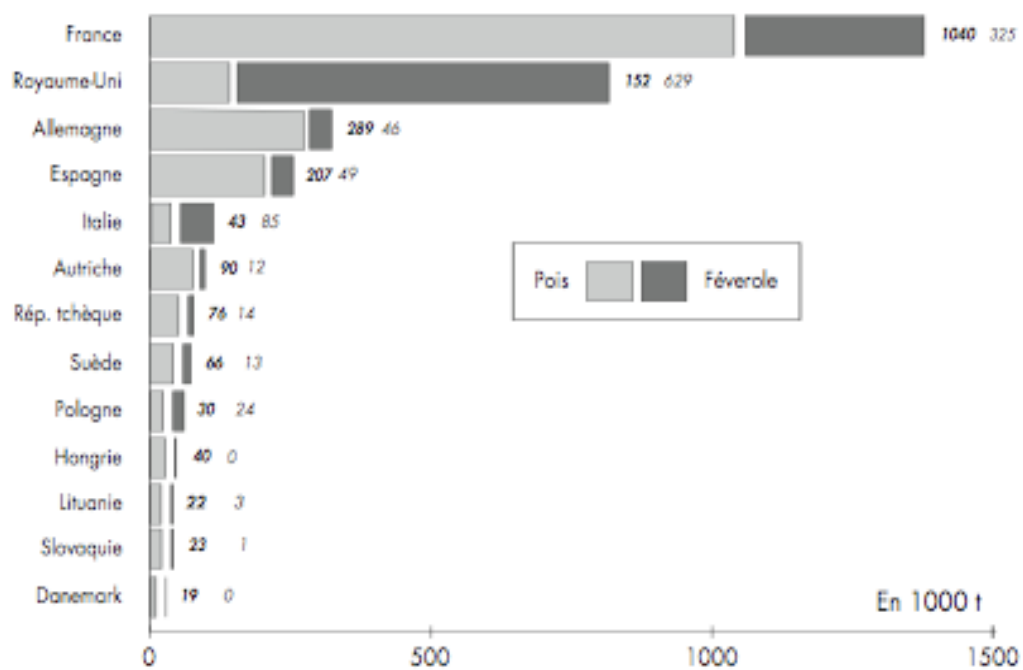
Ce marché est principalement dominé par trois types de semences en fonction de leur contenu: les graines amidonnées (essentiellement les céréales), les oléagineuses et les protéagineuses (essentiellement les graines de légumineuses).

### 1. PRESENTATION DU SECTEUR PROTÉAGINEUX

#### 1.1 Présentation du secteur

##### 1.1.1 Production (données PROLEA, 2005, <http://www.prolea.com/>)

Dans le monde, lorsque l'on considère le secteur protéagineux on parle par ordre croissant de production : du pois sec, du pois chiche, de fève/féverole, de la lentille et du lupin. En Europe, quatre pays se partagent la production de protéagineux, essentiellement le pois et la féverole : avec au premier plan, la France qui représente



Source : UNIP

**Figure I.1 : Production de pois et de féveroles dans les principaux pays producteurs de l'UE en 2006 (données PROLEA)**

plus de 40% de la production de l'Union Européenne, suivie du Royaume-Uni qui a favorisé la production de féverole à celle du pois, puis l'Allemagne et l'Espagne (Figure I.1).

Culturellement, la France a toujours cultivé le pois pour produire des protéines végétales, délaissant actuellement la féverole et le lupin (Figure I.2a, b). Elle est d'ailleurs le premier producteur mondial de pois secs en assurant 25 % de la production mondiale. L'évolution des surfaces cultivées ainsi que le rendement de cette culture ont beaucoup fluctué ces vingt dernières années (Figure I.2c). On a d'abord pu observer une phase d'expansion très rapide jusqu'à 1990, due aux mesures prises dans le plan "protéines" dont l'objectif était de réduire la dépendance protéique de l'Europe suite à l'embargo américain sur le soja en 1973. Puis les surfaces de pois ont atteint un palier en 1993 et ont nettement reculé à partir de cette date. Quant aux rendements, nous pouvons constater une grande instabilité, en effet une hausse significative jusqu'à la fin des années 90 est observée. Cette augmentation des rendements s'explique en grande partie par le progrès de la génétique et par l'accélération du renouvellement variétal. Toutefois, depuis 2000, ces rendements ont souvent été pénalisés par de mauvaises conditions climatiques, la montée en puissance de maladies ainsi que par la migration des cultures vers des terres moins favorables.

La culture du soja n'apparaît pas dans les chiffres précédents, cependant il constitue une source très importante de protéines végétales, connue sous la forme de tourteaux. En effet il occupe une place intermédiaire : le soja appartient aux oléagineux bien que ses graines contiennent une teneur en huile d'environ 20 % et une teneur élevée en protéines (environ 32 %).

### **1.1.2 Consommation : l'Europe, vers une autonomie en protéines ?**

La quantité de protéines végétales consommée par habitant en France a été multipliée par 12 en 40 ans et de telles progressions ont été observées dans la plupart des pays développés (UNIP-ITCF, 1998). Les principales explications de cette augmentation généralisée et extrêmement rapide de la demande en protéines végétales résident dans l'augmentation de la consommation de viande en raison d'une montée du niveau de vie. En effet, le pois est aujourd'hui majoritairement utilisé pour l'alimentation animale. D'ailleurs, en France, les pois sont particulièrement recherchés pour l'alimentation des porcs et des volailles de qualité, pour leur richesse en lysine digestible.

La culture intensive du pois est récente dans la Communauté Européenne. En effet, en 1962, à travers la Politique Agricole Commune (PAC), la Communauté Européenne avait fait le choix de favoriser la production céréalière au détriment des oléo-protéagineux qui seront importés. En 1973, pour des raisons climatiques, la production de soja s'effondra aux Etats-Unis ce qui entraîna la suspension provisoire des exportations. L'embargo américain sur les exportations de soja provoqua une prise de conscience, en France et en Europe, sur notre dépendance excessive vis-à-vis des approvisionnements extérieurs en protéines végétales. Dès 1973, la France a mis en place un plan "protéines" pour favoriser le développement des protéagineux. Ce plan eut pour

(a)

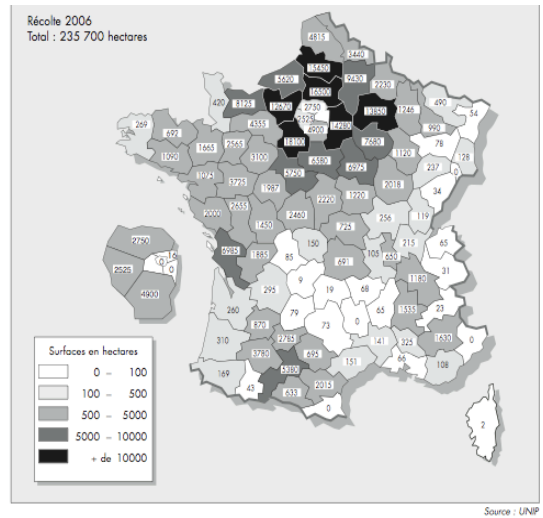
- En année de récolte
- Production et collecte : en 1000 tonnes
- Surfaces : en 1000 ha
- Rendement : en quintaux/ha

Provisoire pour 2006

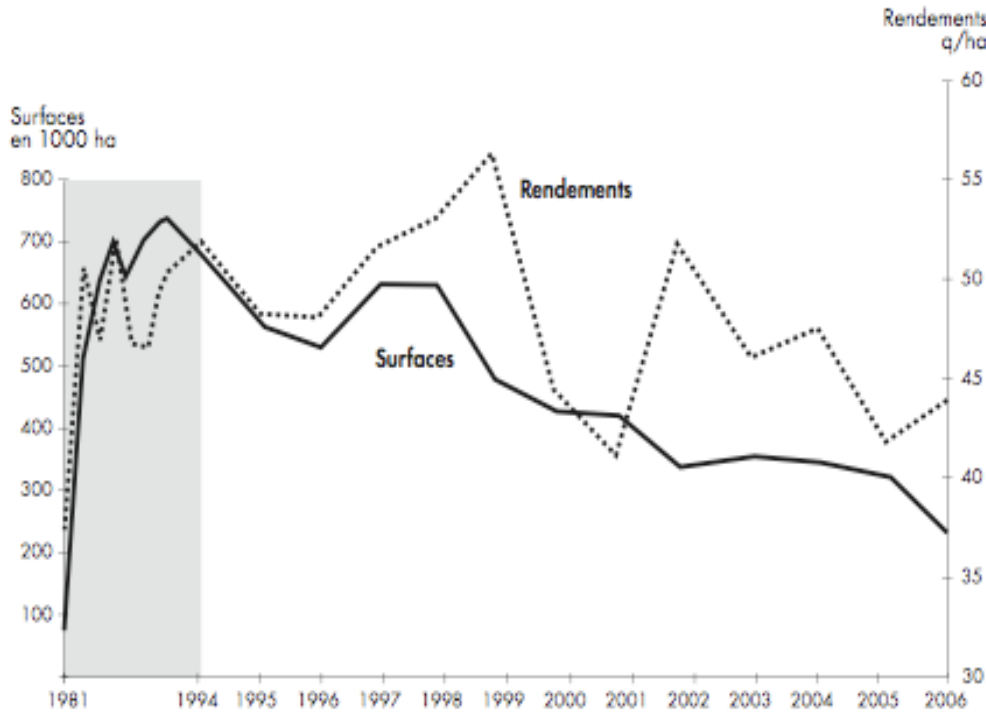
	1981	1993	1998	1999	2000	2001	2002	2003	2004	2005	2006
<b>POIS</b>											
Surfaces	75	737	623	489	424	416	331	363	354	311	236
dont semences certifiées	5	23	25	21	14	18	14,5	14	13,5	11	9
Rendement	38,3	51,1	53,4	56,3	44,5	41,3	51,8	45,9	47,3	41,8	44,2
Production	285	3776	3325	2750	1890	1720	1715	1670	1675	1300	1040
Collecte (1)	260	3360	3000	2400	1652	1346	1365	1385	1395	1150	930
<b>FEVEROLE</b>											
Surfaces	27	13	12,7	13,6	26,4	44	85	80,5	81	102	78
dont semences certifiées	1	0,5	0,5	1	0,9	2,7	3,3	1,9	1,9	2,3	2,3
Rendement	30,3	44,8	38,2	45,0	41,5	38,0	44,8	36,2	46,0	38,9	41,7
Production	81	57	49	61	109	167	381	291	373	397	325
Collecte (1)	55	29	42	55	74	134	270	200	275	305	230
<b>LUPIN</b>											
Surfaces	0	3,5	4,2	5,6	10,9	13,6	15	12,5	9,5	7,4	7
dont semences certifiées	0	0,5	0,4	0,7	1,3	2,3	1,6	0,5	0,7	0,6	0,5
Rendement	-	35	31	35	25	20	28	25	27	24	25
Production	0	12	13	20	27	27	42	31	26	18	18
Collecte (1)	0	6	3	5	9	6	5	5	3	3	3
<b>TOTAL PROTEAGINEUX</b>											
Surfaces	101	754	640	508	461	474	431	456	445	420	321
dont semences certifiées	7	24	26	23	17	23	19	16	16	14	12
Rendement	36	51	53,0	55,7	43,9	40,4	49,6	43,7	46,7	40,8	43,1
Production	366	3845	3387	2831	2026	1914	2138	1992	2074	1715	1383
Collecte	315	3395	3045	2460	1735	1486	1640	1590	1673	1458	1163

(1) non compris les semences et l'auto-consommation.

(b)



(c)



**Figure I.2 : Production de pois en France (données PROLEA)**

- Chiffres concernant les productions de protéagineux en France durant ces dernières années
- Répartition géographique des surfaces cultivées de pois en France en 2006
- Evolution des surfaces et rendements de pois en France

effet de passer à un taux d'autosuffisance de 40% dans les années 80. La production de soja reste cependant faible suite à la signature des accords de Blair House en 1992 qui limitent les surfaces de soja en Europe.

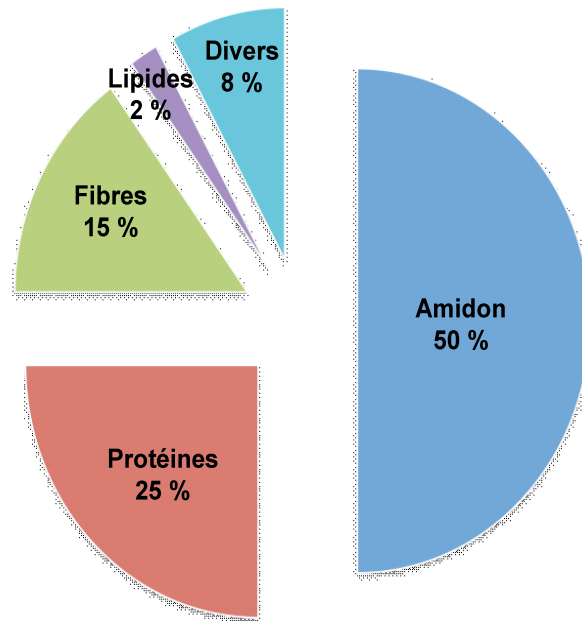
Aujourd'hui, l'Union Européenne est le premier consommateur de tourteaux de soja (plus de 40 millions de tonnes, soit le ¼ de la production mondiale), elle en est également le premier importateur. En conséquence, l'Union Européenne montre toujours une très grande dépendance face aux importations de protéines végétales avec seulement 33% d'auto-provisionnement (Source UNIP 2002/2003). Par conséquent, environ deux tiers de ses besoins en protéines végétales sont encore importés, essentiellement à partir des Etats-Unis.

L'Europe est donc aujourd'hui encore face à un problème et doit s'inscrire dans une démarche volontariste de plus grande autonomie en protéines végétales. En effet, les enjeux politiques et économiques sont très importants : depuis 2000 et l'interdiction des farines animales, le besoin en protéines est plus grand, à ceci s'ajoute une demande mondiale de plus en plus forte des pays émergents. L'embargo de 1973 a montré que la sécurité d'approvisionnement n'était pas garantie d'autant qu'elle est contrôlée par seulement quelques pays (Etats-Unis, Argentine et Brésil). Enfin, il est à noter que ces grands pays producteurs de soja développent des cultures de soja OGM, cultures OGM dont l'importation pourrait se heurter au refus actuel d'une grande partie de la population en France et plus largement en Europe.

Dans ce contexte, l'amélioration du secteur protéagineux est un enjeu majeur pour atteindre une moindre dépendance en protéines végétales. Ainsi, le pois a une place essentielle à tenir, même s'il ne pourra à lui seul combler tout le déficit.

## **1.2 L'amélioration dans le secteur protéagineux : l'origine et les lignes directrices**

Face aux besoins environnementaux, économiques, énergétiques et alimentaires actuels, les protéagineux ont un rôle à jouer pour construire une agriculture durable. Tout d'abord, leurs graines fournissent les protéines végétales indispensables aux hommes et aux animaux. Par ailleurs, la culture des légumineuses a une portée environnementale forte qui s'accorde avec les préoccupations d'une agriculture durable. En effet, leur capacité de fixation de l'azote atmosphérique via une symbiose avec des bactéries du sol (*Rhizobium meliloti*) (Barker *et al.* 1990) permet de réduire significativement l'utilisation d'engrais azotés, dont la fabrication constitue de loin la principale dépense d'énergie fossile en agriculture. A titre d'exemple, 60% du coût énergétique d'une culture de blé est due à la fertilisation azotée (Munier-Jolain et Carrouée, 2005). De plus, cette non-utilisation d'engrais azotés a un effet très significatif sur la production de gaz à effets de serre. Elle réduit en effet l'émission de protoxyde d'azote (N<sub>2</sub>O), qui se dégage lors de l'épandage au champ, le N<sub>2</sub>O qui est un gaz à effet de serre est 300 fois plus puissant que le CO<sub>2</sub>. L'introduction de la culture des légumineuses dans les rotations dominées par les céréales et le colza ont également pour effet de diminuer le développement des mauvaises herbes, des maladies et des ravageurs, fréquemment observés. Enfin, une culture comme celle du pois de printemps requiert moins d'irrigation que les autres cultures d'été en raison de son cycle court.



**Figure I.3 : Composition de la graine de pois (données PROLEA)**

Dans ce contexte, la culture de protéagineux et la production de protéines végétales doivent être développées en Europe. Développement qui doit cibler l'augmentation des surfaces cultivées (seulement 5% de la surface agricole est occupée par des protéagineux en Europe contre 15-25% dans d'autres pays non européens, [www.grainlegumes.com](http://www.grainlegumes.com)) et par l'amélioration du rendement et de la teneur en protéines des graines.

Les différents axes de recherche choisis dans l'amélioration du secteur protéagineux sont :

- Une tolérance accrue de ces plantes aux stress biotiques et abiotiques,
- La conception de systèmes de culture compétitifs et durables en utilisant les avantages environnementaux,
- Une augmentation du rendement en graines,
- Une amélioration de la teneur en protéines des graines de manière quantitative mais également qualitative (amélioration de la digestibilité).

Ces différents axes sont conduits par des instituts de recherche nationaux et internationaux et notamment par l'INRA qui est un acteur majeur dans la recherche pour l'amélioration des protéagineux et particulièrement le pois. Mon travail de thèse, au sein de l'UMR-LEG, s'inscrit dans ce contexte puisqu'il cible l'amélioration de la qualité de la graine de pois en s'appuyant sur *Medicago truncatula*, espèce modèle des légumineuses.

## **2. LES PLANTES D'ETUDE**

### **2.1 Le pois : plante d'intérêt agronomique**

Le pois, *Pisum sativum*, est une plante annuelle de la famille des légumineuses. Il est connu depuis l'Antiquité. Les premières traces de cette plante remontent à trois cents ans avant Jésus Christ, quand Théophraste dans son livre intitulé "Recherches sur les plantes" décrit plusieurs espèces de la famille des légumineuses dont le pois. Au Moyen Age, le pois était considéré comme un légume sec riche en énergie destiné à l'alimentation humaine. Aujourd'hui, cette plante est toujours essentiellement cultivée pour ses graines riches en protéines consommées comme légume (pois potager) ou servant pour l'alimentation animale (pois protéagineux).

En effet, les graines de pois sont caractérisées par une forte teneur en protéines digestibles, en moyenne 25 % de la matière sèche, ce qui est une teneur intermédiaire entre celle du soja (40 %) et celle des céréales (13 %). Les protéines des graines de pois sont les plus concentrées en lysine parmi les matières premières végétales (teneur moyenne en lysine de 7.4 % contre 6.4 % dans le tourteau de soja, [www.prolea.com](http://www.prolea.com)). L'amidon représente presque 50 % des réserves de la graine de pois, qui est également caractérisée par l'absence de tanins et une faible activité anti-trypsique. Le reste de la matière sèche est composé à 2 % de lipides et 15 % de fibres et sucres solubles (Figure I.3).

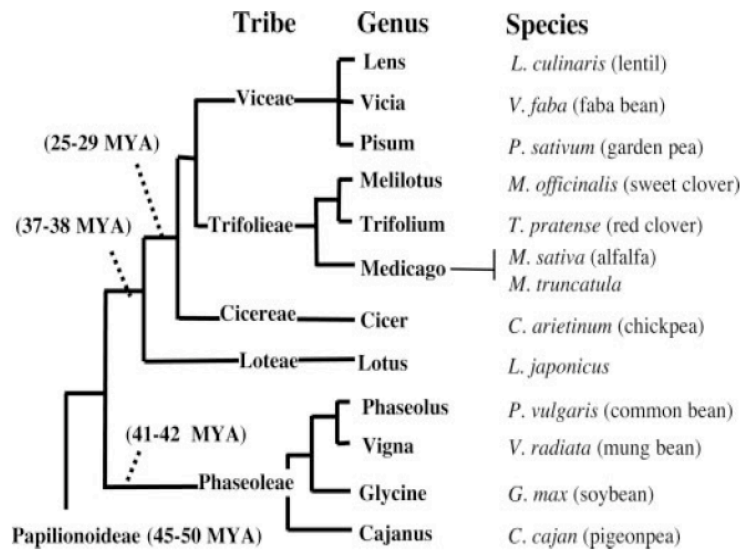


## Estimating genome conservation between crop and model legume species

Hong-Kyu Choi<sup>1,2</sup>, Jeong-Hwan Mun<sup>1,4</sup>, Dong-Jin Kim<sup>1,4</sup>, Hongyan Zhu<sup>1</sup>, Jong-Min Baek<sup>5</sup>, Joanne Mudge<sup>6</sup>, Bruce Roel<sup>1</sup>, Noel Ellis<sup>1,1</sup>, Jeff Doyle<sup>1,1</sup>, Gyorgy B. Kiss<sup>1,5,5</sup>, Nevin D. Young<sup>7</sup>, and Douglas R. Cook<sup>1,5,9</sup>

<sup>1</sup>Department of Plant Pathology and <sup>2</sup>College of Agricultural and Environmental Sciences Genomics Facility, University of California, One Shields Avenue, Davis, CA 95616; <sup>3</sup>Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108; <sup>4</sup>Advanced Center for Genome Technology, University of Oklahoma, Norman, OK 73019; <sup>5</sup>John Innes Centre, Norwich NR4 7UH, United Kingdom; <sup>6</sup>Department of Plant Biology, Cornell University, Ithaca, NY 14853; and <sup>7</sup>Biological Research Center, Institute of Genetics, H-6701 Szeged, Hungary

Edited by Susan R. Wessler, University of Georgia, Athens, GA, and approved August 13, 2004 (received for review March 30, 2004)



**Figure I.4 : Relation phylogénétique entre espèces Légumineuses** (extrait de Choi et al. 2004a)  
L'estimation des temps de divergence sont exprimés en MYA pour "Million Year Ago"

Depuis les premières années de la culture du pois, le rendement a progressé de 0,65 q/ha en moyenne par an, mais demeure cependant variable entre années et zones de culture. Par exemple, en 2004, les rendements ont varié de 30 q/ha à plus de 70 q/ha entre les différentes régions de culture (source Arvalis / UNIP).

La teneur en protéines des graines est une variable importante de la qualité pour les débouchés en alimentation animale. Cependant les fortes fluctuations de cette teneur aussi bien entre années qu'entre zones de cultures gênent les entreprises agro-alimentaires. Par ailleurs, cette teneur reste faible comparée au tourteau de soja et limite le taux d'incorporation du pois dans les aliments. Il est donc nécessaire de comprendre l'origine de la variabilité de la teneur en protéines pour l'améliorer. L'une des premières étapes nécessaires pour améliorer le mécanisme d'accumulation des protéines dans la graine est de l'identifier. En effet, plusieurs études ont été réalisées pour identifier les voies métaboliques impliquées dans le développement de la graine de pois (par exemple Bhattacharyya *et al.* 1990 ; Weber *et al.* 1995 et pour revue Weber *et al.* 2005), en revanche peu d'études concernant la régulation transcriptionnelle du développement de la graine ont été réalisées, en particulier pour la synthèse des protéines de la graine.

## **2.2 *Medicago truncatula* : plante modèle des légumineuses**

Pour élucider les mécanismes moléculaires, une alternative consiste à utiliser des plantes à génome de petite taille. Une telle démarche a été entreprise depuis plusieurs années dans le cadre des travaux réalisés sur la plante modèle crucifère *Arabidopsis thaliana*. Elle a permis des avancées considérables dans la compréhension des bases génétiques et moléculaires régissant le fonctionnement des plantes (Bouchez 2004).

Même si, *A. thaliana* est potentiellement une plante modèle intéressante pour étudier le remplissage de la graine et l'accumulation des protéines de réserve avec une teneur en protéines de 40 % dans la graine mature, on observe d'importantes différences de composition protéique entre cette graine et celles des légumineuses. En effet, la graine mature d'*A. thaliana* est composée d'albumines (2S) et de globulines 12S [seulement un spot protéique de globuline 7S a pu être identifié dans le protéome de la graine mature (Gallardo *et al.* 2001, 2002)]. En revanche, les légumineuses contiennent des albumines 2S, des globulines 11S et une part non négligeable de globulines 7S (Tzitzikas *et al.* 2006). De plus, plusieurs études ont montré une cinétique d'accumulation des protéines de réserve différente chez les légumineuses par rapport à *A. thaliana*. C'est dans ce contexte que l'espèce modèle des légumineuses *Medicago truncatula* (nommée communément luzerne tronquée) (Barker *et al.* 1990 ; Cook 1999, Ané *et al.* 2008) a été préférée à *A. thaliana* pour étudier les mécanismes de remplissage de la graine des légumineuses. En effet, la graine de *M. truncatula* est caractérisée par une teneur en protéines de 32 à 42 % du poids sec, de composition équivalente à celle du pois, de 8-10 % de lipides, de 6-10 % de sucres solubles et moins de 1% d'amidon (Djemel *et al.* 2005). De plus, la cinétique de remplissage est également très proche de celle du pois.

*M. truncatula* est une espèce d'origine méditerranéenne phylogénétiquement proche de la plupart des légumineuses cultivées en Europe (Figure I.4; Choi *et al.* 2004a). Elle présente une forte proximité évolutive et physiologique ainsi qu'une forte synténie avec plusieurs espèces d'intérêt agronomique, parmi lesquelles le pois.



Il s'agit d'une espèce annuelle et diploïde ( $n = 8$ ) au mode de reproduction autogame. Comparativement aux légumineuses cultivées, *M. truncatula* présente un génome de petite taille (environ 500 Mpb), trois à quatre fois supérieur à celui d'*A. thaliana* mais environ neuf fois inférieur à celui du pois.

L'étude de la génétique et de la génomique de *M. truncatula* est aujourd'hui au centre d'un vaste programme de recherche initié à l'échelle mondiale (<http://www.medicago.org>). L'investissement consenti sur cette espèce a déjà permis l'obtention d'un grand nombre de séquences codantes de gènes. Il a également conduit à des avancées rapides dans la construction de cartes génétiques et physiques (Thoquet *et al.* 2002 ; Choi *et al.* 2004b). Des études récentes dans les domaines de la transcriptomique, protéomique et métabolomique ont proposé une caractérisation de l'expression des gènes et du produit de leur expression dans différents organes ou tissus de la plante (Mathesius *et al.* 2001 ; Gallardo *et al.* 2003 ; Watson *et al.* 2003 ; El Yahyaoui *et al.* 2004 ; Buitink *et al.* 2006 ; Gallardo *et al.* 2007). Depuis peu, un consortium entre laboratoires a été constitué pour organiser le séquençage massif du génome de *M. truncatula* (<http://www.medicago.org/genome/>) avec aujourd'hui plus de 2100 BACs séquencés, ce qui correspond à environ 200 Mpb (65%) de l'euchromatine. De plus, de nombreuses populations de mutants (TILLING, fast-neutron, Tnt1) ont été générées sur cette espèce, ce qui offre l'opportunité d'explorer la fonction des gènes ou identifier des gènes essentiels à la plante.

### 3. LE DÉVELOPPEMENT DE LA GRAINE DES LÉGUMINEUSES

De manière classique, le développement de la graine peut être décomposé en trois grandes étapes :

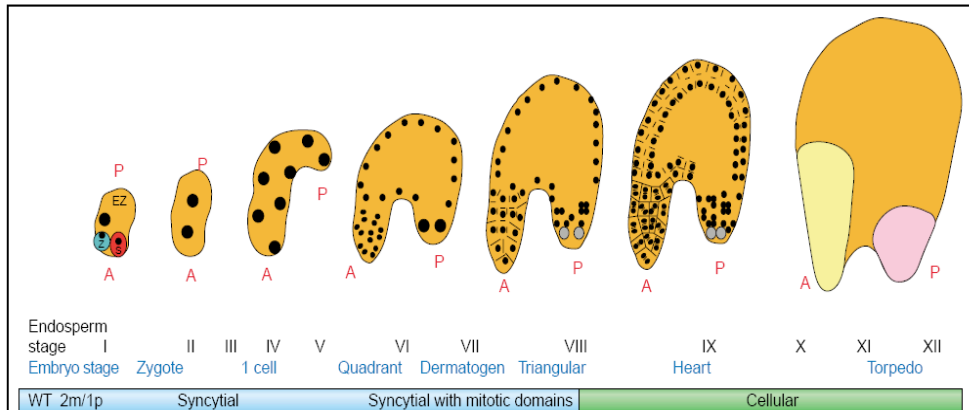
- (i) L'embryogenèse qui est caractérisée par l'acquisition de la polarité ainsi que par la morphogenèse de l'embryon,
- (ii) le remplissage de la graine qui débute avec l'allongement de l'embryon, qui va venir occuper tout l'espace disponible entre les téguments, et qui se caractérise par une accumulation massive des composés de réserve dans les cotylédons,
- (iii) la déshydratation qui correspond au processus d'acquisition de la tolérance à la dessiccation mis en place par la graine en fin de développement, alors qu'elle se déshydrate fortement et rentre en quiescence métabolique.

Ces trois étapes seront décrites plus finement d'un point de vue métabolique dans le chapitre 1, mais au préalable, nous allons évoquer brièvement le processus de développement des différents tissus de la graine : les téguments, l'albumen, et l'embryon, puis la synthèse et l'accumulation des protéines de réserve.

# Endosperm: the crossroad of seed development

Frédéric Berger

Current Opinion in Plant Biology 2003, 6:42-50



**Figure 1.5 : Développement de l'albumen chez *A. thaliana*** (extrait de Berger 2003)

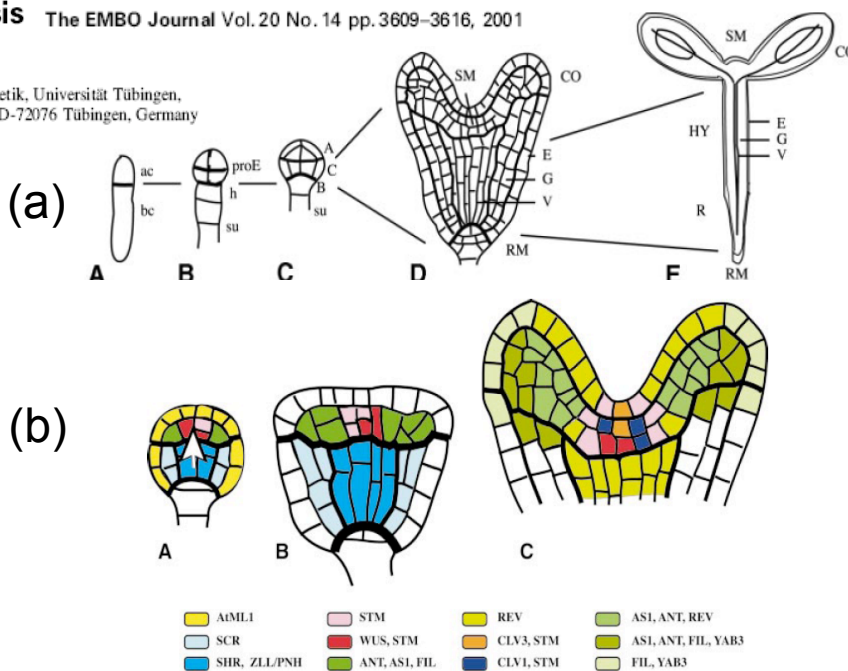
La fécondation du sac embryonnaire produit deux zygotes : le zygote principal (Z) et le zygote accessoire qui donnera l'albumen (EZ). Le développement de l'albumen se divise en deux phases : la formation du syncytium (barre bleue) puis la cellularisation (barre verte). A la fin de la cellularisation, trois domaines sont définis : l'albumen micropylaire (jaune) proche du pôle antérieur (A), l'albumen périphérique (orange) et l'albumen chalazal près du pôle postérieur (P). Les stades de développement de l'embryon correspondant aux différents stades de l'albumen sont indiqués.

## Apical-basal pattern formation in *Arabidopsis*

embryogenesis The EMBO Journal Vol. 20 No. 14 pp. 3609-3616, 2001

Gerd Jürgens

ZMBP, Entwicklungs-genetik, Universität Tübingen,  
Auf der Morgenstelle 1, D-72076 Tübingen, Germany



**Figure 1.6 : Morphogénèse de l'embryon chez *A. thaliana*** (extrait de Jürgens 2001)

(a) Mise en place de la **polarité apicale-basale de l'embryon**. Le zygote subit une première division (A) donnant cellules basales (bc) et apicales (ac). Puis au stade 8 cellules (B), le pro-embryon (proE) dérive de la cellule basale et l'hypophyse (h) et le suspenseur (su) dérivent de la cellule basale. Au stade globulaire (C), trois régions embryonnaires sont distinguables : apicale (a), centrale (c) et basale (b). Au stade cordiforme (D), le plan d'organisation de l'embryon est en place : méristème apical (MA), primordia des cotylédons (CO), méristème racinaire (MR), l'épiderme (E), parenchyme (P) et primordia vasculaire (PV). Le développement de cette embryon aboutira à la plantule (E).

(b) **Origine et organisation du méristème apical**. Le code couleur rend compte du pattern d'expression des gènes régulateurs de l'organogénèse. (A) stade globulaire, (B) stade intermédiaire globulaire-coeur et (C) stade cordiforme.

### 3.1 Formation des trois tissus de la graine

Suite à la double fécondation du sac embryonnaire, nous allons obtenir : le zygote principal (2n), issu de la fécondation de l'oosphère, qui va se diviser activement pour former l'embryon et le zygote accessoire (3n), issu de la fécondation de la cellule centrale, qui formera l'albumen. Ces deux tissus sont entourés des tissus diploïdes maternels, les téguments.

Les téguments sont composés d'assises d'origine maternelle situées en périphérie de la graine et qui se divisent en trois tissus : le tégument externe, le tégument interne et l'endothélium. L'endothélium qui est l'assise la plus interne, est le siège de la synthèse des pigments qui donneront sa couleur à la graine mature chez *A. thaliana*.

L'albumen est issu du zygote accessoire, le premier noyau se divise de manière répétée jusqu'à la formation d'un albumen syncytial (Mansfield et Briarty, 1990a). Les noyaux s'accumulent à la périphérie du sac embryonnaire et progressivement, ce syncytium s'organise en domaines cytoplasmiques nucléaires (Figure I.5, Berger 2003). La cellularisation de l'albumen débute dans la région entourant l'embryon, au moment où ce dernier initie ses cotylédons. Alors que la taille de l'embryon augmente, l'albumen va peu à peu disparaître. Durant la période de maturation, cet albumen est réduit à une assise cellulaire périphérique chez *M. truncatula*, tandis qu'il disparaît totalement chez le pois (Djemel *et al.* 2005).

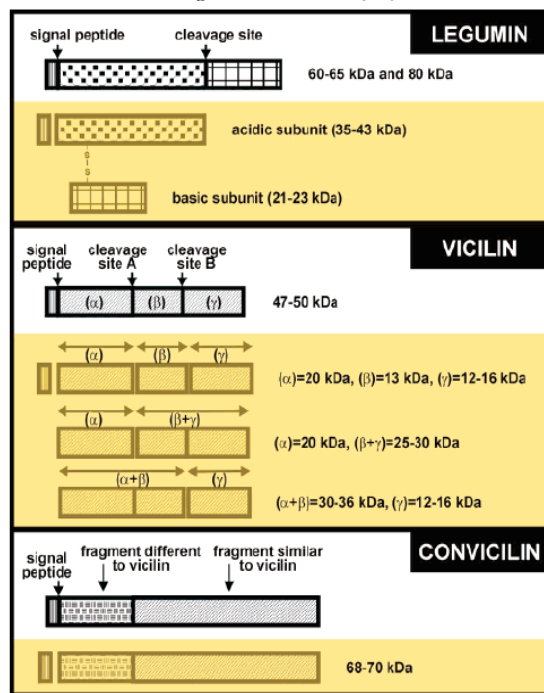
L'embryon est directement issu du zygote principal et va subir une première division asymétrique générant les cellules basales et apicales. La cellule apicale, de petite taille et riche en cytoplasme, va donner naissance à des cellules pro-embryonnaires par le jeu de divisions verticales et horizontales. La cellule basale, de plus grande taille, contient une vacuole et se divise horizontalement, de manière répétée, engendrant un groupe de quelques cellules (Figure I.6a, Jürgens, 2001). Ce groupe cellulaire formera le suspenseur à l'exception de la cellule directement adjacente aux cellules pro-embryonnaires qui fera partie intégrante de l'embryon et donnera naissance au méristème racinaire (Jürgens, 2001, pour revue). Ces premières étapes de divisions cellulaires essentielles pour l'histodifférenciation de l'embryon sont régulées par le facteur de transcription *MERISTEM LAYER 1* qui est exprimé spécifiquement dans la cellule apicale. Cette polarité apicale-basale ainsi acquise donnera lieu à l'élaboration de différentes régions distinguables dès le stade torpille : une région apicale (méristème apical et cotylédons), un hypocotyle et une racine incluant un méristème racinaire (Mayer *et al.* 1991). De plus, d'autres niveaux d'organisation viennent progressivement s'y superposer avec la mise en place d'une organisation radiale qui sera particulièrement visible dans l'hypocotyle, et qui implique trois identités tissulaires distinctes, avec, depuis la périphérie vers le coeur de l'embryon, l'épiderme, le tissu intermédiaire parenchymateux, et le procambium central.

Chacune de ces régions va ensuite évoluer de manière autonome sous le contrôle de l'expression différentielle de nombreux gènes dont des facteurs de transcription (pour revue Mayer *et al.*, 1991 et Jürgens 2001). Un exemple de ce contrôle exercé par les facteurs de transcription sur le développement de l'embryon est visible lors de la mise en place des méristèmes racinaire et apical.

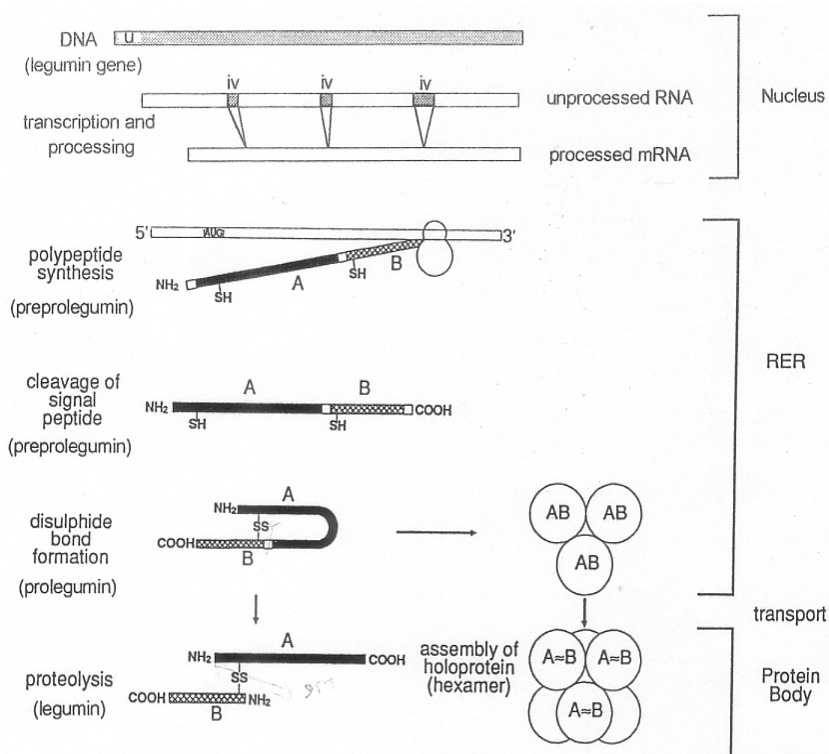
Le méristème racinaire (RAM), à l'extrémité basale de l'embryon, présente un centre quiescent, formé de

## Genetic Variation in Pea Seed Globulin Composition

EMMANOUIL N. TZITZIKAS,<sup>†,§</sup> JEAN-PAUL VINCKEN,<sup>\*,†</sup> JOLAN DE GROOT,<sup>#,1</sup>  
 HARRY GRUPPEN,<sup>#</sup> AND RICHARD G. F. VISSER<sup>†</sup> *J. Agric. Food Chem.* 2006, 54, 425–433



**Figure 1.7 :** Description schématique de la famille des globulines (extrait de Tzitzikas *et al.* 2006). Représentation sur fond blanc de la protéine sous forme précurseur et sur fond jaune de(s) formes(s) matures. Les poids moléculaires et les sites de clivage sont indiqués.



**Figure 1.8 :** Synthèse des légumines dans les cotylédons (extrait de Seeds, Bewley et Black, 1994).

Les différentes étapes de la synthèse de légumine sont représentées dans ce schéma et décrites dans le texte. Les différents compartiments cellulaires sont mentionnés. L'abréviation A correspond à la chaîne acide et B à la chaîne basique.



quatre cellules mitotiquement inactives qui confèrent aux cellules avoisinantes une identité de cellule souche en empêchant leur différenciation (Jürgens, 2001). Cette zone de croissance sera à l'origine de tous les tissus différenciés de la racine.

Le méristème apical (SAM) commence sa formation dès le stade globulaire précoce (Figure I.6b ; Jürgens, 2001, pour revue). Un des premiers gènes à être exprimé, est le facteur de transcription *WUSCHEL (WUS)*. Son domaine d'expression est restreint à la région interne du méristème et il accompagne la formation de ce dernier. L'expression des autres facteurs de transcription *SHOOTMERISTEMLESS (STM)*, *CUPSHAPED COTYLEDON 1 et 2 (CUC1, CUC2)*, et *ANTEGUMENTA (ANT)* est initiée après celle du gène *WUS*, dans l'embryon globulaire tardif, et leur expression permet d'initier la mise en place progressive des identités cellulaires dans la partie caulinaire de l'embryon et ainsi aboutir à une symétrie bilatérale (Figure I.6b ; Jürgens, 2001) : une région centrale exprimant *STM* et *CUC2* donnera naissance au SAM, deux zones périphériques opposées exprimant *ANT* donneront naissance aux cotylédons et deux zones périphériques alternées avec les précédentes et exprimant *STM*, *CUC2* et *ANT* formeront les frontières entre les cotylédons.

### **3.2 Synthèse et accumulation des protéines de réserve**

Suite à ces étapes d'histodifférenciation, l'embryon au stade torpille a acquis sa polarité. Chez *M. truncatula*, on considère que ce stade est atteint à 12-14 Jours Après Pollinisation (JAP). A ce stade, l'embryon qui a principalement été soumis à une intense activité de division cellulaire va basculer dans la phase de remplissage. Sa taille va augmenter sous l'effet d'une forte expansion cellulaire qui est accompagnée d'une importante accumulation de molécules de réserve dont des protéines.

La synthèse et l'accumulation de ces protéines sont soumises à une régulation spatio-temporelle très stricte. Ces processus ont lieu dans les cotylédons et plus particulièrement dans des organites de stockage, les corps protéiques. Chez les légumineuses, ces corps protéiques sont des vacuoles de stockage qui servent de compartiments de stockage des molécules de réserve pendant le remplissage et qui serviront également de site d'hydrolyse des macromolécules lors de la phase pré-germinative. Chez les légumineuses, les protéines de réserve sont de deux types : les albumines (2S) qui sont généralement minoritaires et les globulines qui représentent environ 70 % de l'azote total contenu dans les graines. Ces globulines sont constituées de deux familles de protéines qui sont identifiables par leur masse moléculaire et donc par leur coefficient de sédimentation. On distingue donc la famille des 7S composée de vicilines et convicilines et la famille des 11S composée des légumines (Figure I.7). Les protéines de réserve 11S sont stockées sous forme d'un complexe hexamérique de 320-400 kDa composé de six sous-unités différentes (52-65 kDa). Chaque sous-unités est elle-même composée d'un polypeptide acide (appelé chaîne acide d'un pI de 6,5 et de 33-42 kDa) et d'un polypeptide basique (appelé chaîne basique d'un pI de 9 et de 19-23 kDa). Ces deux chaînes sont liées par un pont disulfure. Bien que l'accumulation des protéines de réserve ait lieu dans les vacuoles de stockage, leur synthèse a lieu dans le réticulum endoplasmique rugueux (RER) puis elles sont transportées dans les vacuoles de stockage. En effet, les chaînes acides et basiques sont initialement issues du même ARNm. Le produit de transcription puis de traduction contient donc ces deux chaînes mais également un peptide signal (cette forme est appelée



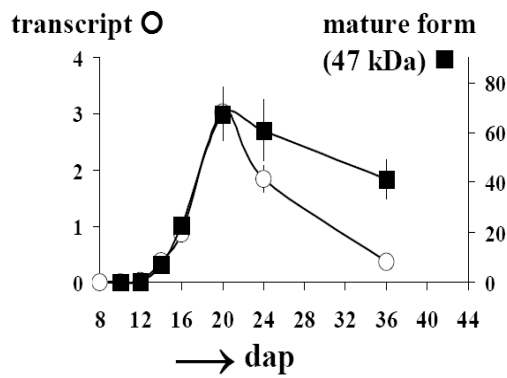
## A Combined Proteome and Transcriptome Analysis of Developing *Medicago truncatula* Seeds

EVIDENCE FOR METABOLIC SPECIALIZATION OF MATERNAL AND FILIAL TISSUES<sup>13</sup>

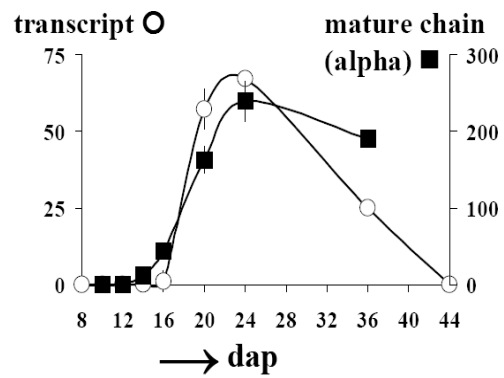
Karine Gallardo<sup>‡§</sup>, Christian Firnhaber<sup>¶</sup>, Hélène Zuber<sup>‡</sup>, Delphine Hélicher<sup>‡</sup>, Maya Belghazi<sup>||\*</sup>, Céline Henry<sup>‡‡</sup>, Helge Küster<sup>¶</sup>, and Richard Thompson<sup>‡</sup>

# B

### Vicilin relative abundance



### Legumin B relative abundance



**Figure I.9:** Contrôle transcriptionnel de la synthèse des vicilines et des légumines B chez *M. truncatula* (extrait de Gallardo et al. 2007).

Les graphiques représentent les cinétiques d'accumulation des transcrits par rapport à celles des protéines durant le développement de la graine (de 8 à 44 jours après pollinisation).

préprolégumine). Le peptide signal est clivé et la protéine acquiert sa structure tertiaire et quaternaire (prolégumine). Ce complexe de prolégumines est transloqué dans les vacuoles de stockage via l'appareil de Golgi. C'est une fois dans ces organites de stockage que les deux chaînes acides et basiques seront clivées puis assemblées en hexamères pour former des légumines (Figure I.8).

Par rapport aux protéines 11S, les protéines 7S sont stockées sous forme de complexes trimériques de 145-190 kDa, composées de trois polypeptides différents de 48 à 83 kDa. Ces polypeptides sont formés à base de précurseurs qui subissent une série de sites de protéolyse pour aboutir à des polypeptides allant de 12 à 75 kDa.

#### **4. CONTRÔLE TRANSCRIPTIONNEL DE LA SYNTHÈSE DES PROTÉINES DE RÉSERVE**

La régulation transcriptionnelle, et donc les facteurs de transcription, joue un rôle majeur dans l'étape d'histodifférenciation de l'embryon, décrite précédemment (section 3.1). Il en est de même pour la régulation de la phase de remplissage avec la synthèse des protéines de réserve qui est principalement régulée de manière transcriptionnelle chez les légumineuses (Meinke 1981 ; Gatehouse et al. 1982 ; Bassuner et al 1983 ; Walling et al 1986 ; Boulter 1987) (Figure I.9). En effet, ces gènes contrôlent de manière spatio-temporelle très stricte la synthèse des protéines de réserve. Pour ce travail de thèse, nous nous sommes particulièrement intéressés aux gènes pouvant intervenir dans cette régulation chez les légumineuses. Cependant, les données concernant le contrôle transcriptionnel du remplissage de la graine se sont révélées éparpillées car provenant de nombreuses études réalisées ces dix dernières années sur *P. vulgaris* ou *A. thaliana*. Par conséquent, nous nous sommes proposés de regrouper ces résultats afin d'obtenir une vision globale des mécanismes transcriptionnels impliqués lors de cette phase critique pour la qualité du grain. Cette partie se réfère à une revue acceptée dans le journal *Plant and Cell Physiology*.



## **Transcriptional regulation of storage proteins during dicotyledon seed filling**

**Jérôme Verdier** and Richard D. Thompson

Revue acceptée dans *Plant and Cell Physiology* (modifications mineures - Juin 2008)

### Résumé :

Les graines représentent une importante source de nutriment pour l'alimentation humaine et animale. La valeur nutritive des graines est en grande partie due aux molécules de réserve qui s'accumulent au cours de la phase de remplissage de la graine. Au cours de ces dernières années, notre compréhension des mécanismes de régulation du remplissage de la graine a progressé de manière significative grâce à la diversité des approches expérimentales utilisées. Cette revue synthétise les récents travaux relatifs aux facteurs de transcription qui régulent l'accumulation de protéines de réserve. Un modèle de régulation de la synthèse de ces protéines est établi décrivant les événements avant, pendant et après la synthèse des protéines de réserve. Le contrôle transcriptionnel de la régulation des protéines de réserve est accompagné par un contrôle environnemental et physiologique, notamment à travers l'action des hormones et des métabolites. Enfin, de récentes analyses post-génomiques sur différentes plantes modèles ont établi l'existence d'un mécanisme de remplissage de la graine conservé impliquant le rôle des gènes régulateurs maîtres du développement (*LEC1*, *LEC2*, *ABI3* and *FUS3*), alors que d'autres analyses ont révélé des différences dans la régulation fine entre les différentes plantes.

Mots clés : Dicotylédones; Protéines de réserve des graines; Facteurs de transcription.



**Title:** Transcriptional regulation of storage proteins during dicotyledon seed filling

**Authors:** Jérôme Verdier and Richard D. Thompson

Unité Mixte de Recherche en Génétique et Ecophysiologie des Légumineuses à Graines (UMR-LEG). Institut National de la Recherche Agronomique (INRA), BP 86510, F-21065 DIJON France.

**Abbreviations:**

ABA response element, ABRE; Abscisic acid, ABA; ABSCISIC ACID-INSENSITIVE3, ABI3; ABSCISIC ACID-INSENSITIVE5, ABI5; Amplified Fragment Length Polymorphism, AFLP, Days after pollination, DAP, FUSCA3, FUS3; Gibberellic acid, GA ; Heme activated protein, HAP ; LEAFY COTYLEDON1, LEC1; LEAFY COTYLEDON2, LEC2; *P. vulgaris* ABI3-like factor, PvALF ; Quantitative reverse transcription PCR, qRT-PCR; Repressor of maturation gene, ROM ; Seed storage protein, SSP; Transcription factor, TF.



## **ABSTRACT**

Seeds represent a major source of nutrients for human and animal livestock diets. The nutritive value of seeds is largely due to storage products which accumulate during a key phase of seed development, seed filling. In recent years, our understanding of the mechanisms regulating seed filling has advanced significantly due to the diversity of experimental approaches used. This review summarizes recent findings related to transcription factors that regulate seed storage protein accumulation. A framework for the regulation of storage protein synthesis is established which incorporates the events before, during and after seed storage protein synthesis. The transcriptional control of storage protein synthesis is accompanied by physiological and environmental controls, notably through the action of plant hormones and other intermediary metabolites. Finally, recent post-genomics analyses on different model plants have established the existence of a conserved seed maturation process involving the master regulators (LEC1, LEC2, ABI3 and FUS3) but also revealed certain differences in fine regulation between plant families.

Keywords: Dicotyledons; Seed storage proteins; Transcription factors.





## **INTRODUCTION**

### ○ **Pre-eminence of transcriptional regulation in seed maturation**

Eukaryotes employ diverse mechanisms to regulate gene expression and activity, including chromatin condensation, DNA methylation, transcriptional regulation, alternative splicing of RNA, mRNA stability, translational control, post-translational modification and protein degradation. Since the early 70's, the theory of Britten and Davidson (1969) has provided the basis of the gene specific transcriptional regulation. This mechanism is now recognized to be the most common point controlling the rate of transcriptional initiation, and appears to be a major determinant of the overall gene expression profile. In plants, gene regulation at the level of transcription is fundamental in developmental processes (Riechmann and Ratcliffe 2000), cell cycle regulation (Huntley and Murray, 1999), stress (Singh et al. 2002) and hormonal (Sun and Gubler, 2004; Busk and Pagès, 1998; McCourt, 1999) responses. Transcription factors (TFs), account for nearly 6% of all plant genes in the *Arabidopsis* genome (Riechmann and Ratcliffe 2000), although only 10% of these have been genetically and functionally characterized (Qu and Zhu, 2006). The *Arabidopsis* TF complement includes at least eight seed-preferentially expressed factors, the four “master regulators”: LEC1, LEC2, ABI3, and FUS3; and in addition EEL, bZIP67 (DPBF2), PEI1, and HSFA9 (an ABI3-regulated heat shock factor).

During seed maturation, the embryo enlarges to accommodate the deposition of storage compounds, which are required for germination and which represent the major economic interest of grain. Carbon is stored in the form of carbohydrates, lipids and storage proteins, the latter enabling in addition the storage of nitrogen. In contrast to monocotyledons, which mainly accumulate seed storage proteins (SSP) in the endosperm (Shotwell and Larkins, 1989), dicotyledonous storage proteins, which consist of 7S and 11S globulin classes, are deposited in the cotyledons (Meinke et al 1981; Walling et al 1986; Ladin et al 1987; Pang et al 1988), although this regional specificity may not be absolute (Greenwood and Chrispeels, 1985; Bäumllein et al. 1991). In vitro transcription assays (Evans et al 1984; Beach et al 1985; Walling et al 1986; Harada et al 1989 and Nielsen et al 1989) revealed that seed storage protein expression is primarily controlled by transcriptional regulation during the maturation phase, the timing of accumulation of the proteins being closely related to that of the corresponding mRNAs (Meinke et al. 1981; Gatehouse et al. 1982; Bassuner et al 1983; Walling et al 1986).

### ○ **The roles of genetically characterized transcription factors (TFs)**

Certain TFs have been independently identified by their mutant phenotypes as playing key roles in different steps of seed development. These include regulators of the maternal contribution to seed size (TTG2, Garcia et al. 2005), mutants of embryo development (MONOPTEROS, Hardtke et al. 1998), and factors affecting the maturation phase and the subsequent phase of desiccation tolerance, such as LEAFY COTYLEDON1 (Kagaya et al. 2005b), LEAFY COTYLEDON2 (Santos Mendoza et al. 2005), FUSCA3 (Bäumllein et al., 1994, Keith et al ., 1994), ABSCISIC ACID-INSENSITIVE3; (Parcy et al. 1997), and



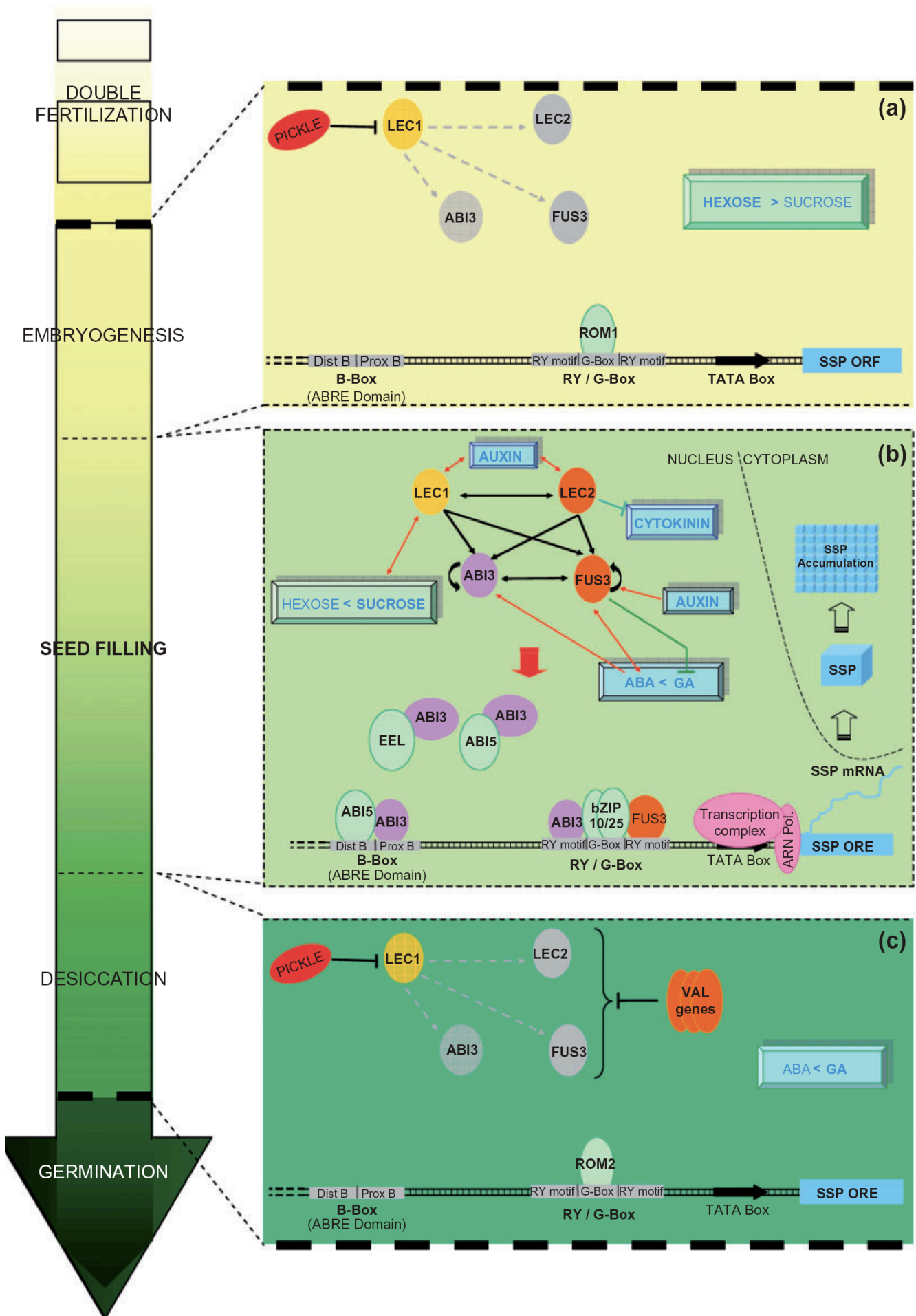
WRINKLED1 (Cernac et al. 2004), Numerous studies have revealed interactions between different regulator genes and cis-regulatory elements.

In this mini-review, we will focus on the TFs implicated in seed storage protein synthesis during dicotyledonous embryo development. According to the functional and biochemical characterisations of these, we propose a detailed model of the gene specific regulation of storage proteins in dicotyledons. The non-genetic control of seed maturation will be also briefly summarized to provide an overall view of this process. Finally, perspectives on the contribution of global genomics approaches will be discussed.

## **TRANSCRIPTIONAL REGULATION OF SEED STORAGE PROTEINS**

### **1/ The cis-regulatory elements identified on seed storage protein genes**

The regulatory cis-elements implicated in promoter activation during the seed maturation phase have been investigated. Extensive analyses were performed on the napin A promoter from *B. napus* and on the legumin B4 promoter from *V. faba* (Stalberg et al. 1996; Ezcurra et al. 1999; Bäumlein et al. 1992). Sequence comparison of the seed storage protein promoters in combination with DNA-protein binding assays revealed two major conserved factor-binding sites, the RY/G motif and the B-box, which act in synergy. Two further cis-motifs (AACA motif, CTTT motif), important in monocot seed gene expression (Vicente-Carbajosa and Carbonero (2005)), will not be treated here as their role in dicot seed expression is not yet established. RY/G motifs are composed of two RY elements (CATGCA), with which B3 domain proteins interact (Bäumlein et al. 1986; Reidt et al. 2000; Ezcurra et al. 2000), and a G-box, (CACGTG), which binds bZIP or bHLH transcription factors (De Pater et al. 1993; Kawagoe et al. 1996), and is responsible for the ABA-response (Chandrasekharan et al., 2003). The B-box is composed of DistB (GCCACTTGTC) and ProxB (CAAACACC) elements and mediates a strong ABA-response in seeds (Ezcurra et al. 1999). The B-box constitutes an ABA response element (ABRE) domain (Ezcurra et al. 1999), whose elements interact with various factors like bZIP or MYB (Ezcurra et al 2000; Pla et al 1993; Hobo et al 1999). In vitro binding assays have shown that seed-specificity is conferred *via* the combinatorial interaction between these boxes (Ezcurra et al. 1999). Moreover, mutations in either of these elements strongly reduce promoter activity (Kawagoe et al. 1994; Ezcurra et al. 1999; Reidt et al, 2000, Chandrasekharan et al. 2003). By site-directed mutagenesis in 10 conserved motifs within 295 bp of the *phas* promoter, Chandrasekharan et al. also identified certain RY motifs as contributing to promoter activity in specific regions of the embryo, the radicle or the hypocotyl. By combining the available information, a summary of events prior to, during and at the end of, storage protein accumulation in the maturing embryo is proposed (Figure 1).



**Figure 1: Summary of events occurring during the gene specific transcriptional regulation of seed storage proteins.** Seed development is represented in three stages corresponding to (a) late embryogenesis, (b) seed filling and (c) desiccation.

## **2/ Prior to seed storage protein synthesis**

Chern et al. (1996a) identified an embryo-specific bZIP factor named ROM1, (Repressor Of Maturation gene1), which interacts with the RY-G motif at the G-Box. ROM1 inhibited enhancer-dependent activation of gene expression in presence of the transcriptional activator PvALF (*P. vulgaris* ABI3-like factors) in bean cotyledons. ROM1 mRNA is present during early embryogenesis but is down-regulated at the beginning of seed maturation, suggesting a possible repressor role of maturation gene transcription, effective until the expression of a maturation signal in embryos, or possibly repressing maturation phase gene expression in non-storage cell-types of the embryo.

## **3/ Entry into seed maturation**

To identify genes involved in seed maturation, plants defective in storage protein accumulation, acquisition of dormancy or desiccation tolerance have been analyzed, resulting in the identification of the *ABSCISIC ACID INSENSITIVE3* (*ABI3*) and *FUSCA3* (*FUS3*) loci. These genes, which belong to the plant-exclusive B3 domain family, are specifically expressed in seeds during the maturation phase. *ABI3* is an orthologue of *Zea mays* *VIVIPAROUS1* (*VPI*) (Mc Carty et al. 1991; Giraudat et al. 1992; Parcy et al. 1995; Bobb et al. 1995), a key player in the regulation of maize kernel maturation, which is ABA-dependent. *fus3* was one of mutant loci causing anthocyanin accumulation in the developing embryo (Castle et al. 1994; Baumlein et al. 1994; Keith et al. 1994). *FUS3* was subsequently shown to be necessary for the specification of cotyledon identity; mutants display leaf characteristics such as trichomes, desiccation-sensitivity, and vivipary, and fail to accumulate storage proteins. *FUS3* was shown to be a direct positive regulator of maturation gene expression (Reidt et al. 2000; Kagaya et al. 2005a) and also to negatively regulate *TRANSPARENT TESTA GLABRA 1* (*TTG1*), which specifies epidermal features of the cotyledon (Tsuchiya et al. 2004).

An ABA-insensitive mutant, *abi3*, shares with *fus3* the phenotypes of desiccation-sensitivity and reduction in seed storage protein accumulation. Even if *ABI3* and *FUS3* appear to have similar action, gene activation by *ABI3* differs to that by *FUS3* in its kinetics, the response to *FUS3* being slower, suggesting the requirement for synthesis or turnover of an intermediate factor (Kagaya et al. 2005a). Both *FUS3* and *ABI3* proteins are able to interact with the RY motif on target promoters thanks to their B3 domains (Mönke et al. 2004; Reidt et al. 2000 and 2001; Bobb et al. 1997; Ezcurra et al. 2000). An activation domain is also identified in both genes that is capable of inducing gene expression (Mönke et al. 2004). Whereas *FUS3* was able to bind the RY motif directly using in vitro and in vivo experiments (Kroj et al. 2003), direct binding of *ABI3* was only observed in vitro (Mönke et al. 2004) suggesting the possibility of association via a second factor to enhance its binding (Kroj et al, 2003).

Lara et al. (2003) have identified two bZIPs, *AtbZIP10* and *AtbZIP25*, which are co-expressed during seed maturation with Arabidopsis seed storage proteins but are also expressed, at higher levels, in vegetative tissues. These bZIPs bind to the G-Box and both interact with the B2 domain of *ABI3* in two-hybrid assays. Transient expression assays showed a synergistic effect of these bZIPs with *ABI3* on seed storage protein gene

**Figure 1: Summary of events occurring during the gene specific transcriptional regulation of seed storage proteins.** Seed development is represented in three stages corresponding to (a) late embryogenesis, (b) seed filling and (c) desiccation.

Genetic and molecular interactions in the regulatory network of control of seed storage protein synthesis during these three stages are indicated. Arrows and T-bars reflect respectively a positive or negative effect. The four master regulatory genes (*LEC1*, *LEC2*, *ABI3* and *FUS3*) are represented at each stage with black arrows and black T-bars when active and with dotted grey arrows when not expressed due to repression of *LEC1* gene expression. Hormonal and metabolic states are indicated for each stage with green and blue boxes respectively. Red arrows and green T-Bars correspond respectively to positive or negative effects. TF families are represented using different colours: yellow for CBF (CCAAT Binding Factor)-type factors, orange for B3 domain-type, green for bZIP-type, and purple for VP1/ABI3 factors.

Cis-regulatory DNA sequence elements present on seed storage protein promoters are indicated: the B-box with DistB (GCCACTTGTC) and ProxB (CAAACACC) elements and the RY/G Box with RY motif (CATGCA) and a G-box (CACGTG). Potential interactions between transcription factors and motifs involved in SSP transcription are indicated at each stage.

- (a) During embryogenesis, *LEC1* expression is repressed by PICKLE and SSP expression is repressed by ROM1 which binds G-Box motifs. Hexose concentration is higher than sucrose concentration.
- (b) During seed filling, sucrose concentration becomes higher than hexose concentration, and *LEC1* is no longer repressed by PICKLE, permitting activation of the four master genes of seed development and resulting SSP gene transcription.
- (c) During desiccation, the master regulatory genes are repressed by PICKLE and three VAL genes, and the GA/ABA ratio increases. SSP transcription is repressed by ROM2 which binds the G-box motif.

List of abbreviations: ABRE, ABA response element; RNA pol, RNA polymerase; SSP, Seed Storage Protein; SSP ORF, Seed storage protein open reading frame VAL genes, VP1/ABI3-LIKE genes.

activation (Lara et al. 2003). Thus, whereas FUS3 binds directly to the RY motifs, ABI3 may need the synergistic effect of two bZIPs binding to the adjacent G-Box in order to enhance its activation binding. In support of this hypothesis, a bZIP factor, TRAB1, has been shown to mediate ABA-induced transcription at ABREs via its interaction with VP1, a rice ABI3-like factor, in plant cells (Hobo et al. 1999). This interaction between ABRE and ABI3 factors has also been shown in *B. napus* on the distB motif of the napin promoter B-Box (Ezcurra et al. 2000) suggesting its similar action with a bZIP factor. Two bZIP factors have been identified during the maturation phase, ABI5 (ABSCISIC ACID INSENSITIVE 5) and EEL (ENHANCED EM LEVEL). ABI5 is expressed in the late maturation phase, and is able to bind the ABRE motif, activating ABA-mediated transcription (Nakamura et al. 2001). It physically interacts with ABI3 (Nakamura et al. 2001), which is consistent with the properties of its putative rice homologue TRAB1. EEL is expressed at the mid-maturation stage and modulates the ABI5-mediated activation, as it competes for the same binding sites and represses ABI5-regulated genes (Bensmihen et al. 2002).

The Arabidopsis mutants, *leafy cotyledon1 (lec1)* and *leafy cotyledon2 (lec2)*, encode further TFs needed to specify cotyledon identity. These mutants share with *fus3* and *abi3* certain traits such as cotyledons partially transformed into leaves, viviparous embryos (Meinke et al. 1994), reduced desiccation tolerance (Meinke 1992; Bäumllein et al 1994; Keith et al 1994; Meinke et al 1994; West et al 1994) and defective synthesis and accumulation of specific maturation proteins including seed storage proteins. LEC1 and LEC2 are required for the maintenance of cotyledon identity throughout maturation, and for the completion of the maturation phase (Braybrook et al. 2006). LEC1, LEC2, FUS3, and ABI3 have been termed “master regulators” of seed maturation, due to their pleiotropic, overlapping and partially redundant functions (Kroj et al., 2003, Santos-Mendoza et al., 2008). Remarkably, ectopic expression of either LEC1 or LEC2 induces the embryogenic pathway in vegetative cells (Santos-Mendoza et al. 2005; Lothan et al. 1998; Alemanno et al. 2008).

*LEC1* encodes a protein with sequence similarity to the HAP3 subunit of CCAAT binding factors (CBF or NF-Y factors, Edwards et al. 1998; Maity and De Crombrugge, 1998). In plants, CBF is a heterotrimer composed of HAP2, HAP3 and HAP5 subunits. In contrast to yeast, multiple isoforms of each subunit have been found (Edwards et al. 1998). In Arabidopsis, the HAP3 subunit genes can be divided into the LEC1-type HAP3 subunits and the non LEC1-type. LEC1 and non LEC1-types differ by 16 amino acid residues that serve as signatures of their conserved domains (Kwong et al. 2003). Ectopic expression of LEC1 induces expression of *FUS3*, *ABI3* and *LEC2*, implying a hierarchical activation network (Kagaya et al. 2005b). LEC2 is a member of the B3 domain-containing TF family, closely related to FUS3 and ABI3. It is also capable of activating *FUS3* and *ABI3*, and of directly binding RY motifs to activate seed storage protein genes (Kroj et al 2003; Mönke et al. 2004; Braybrook et al. 2006). Moreover, *LEC1* and *LEC2* genes positively regulate themselves and each other, forming a regulatory feedback essential for their sustained and cell type-specific expression (To et al. 2006). An additional level of complexity in the interaction network is reflected by differences in their spatial localization within the embryo (To et al. 2006), resulting in differing genetic controls in the cotyledon, hypocotyl and radicle cells (Santos-Mendoza et al., 2008).





Hall's group has investigated the chromatin remodeling underlying the stepwise activation of the phaseolin storage protein promoter (Li et al 2001). Recently, they inducibly expressed the PvALFs, FUS3, and ABI5, ectopically in leaves to monitor their effects in trans on a phaseolin reporter construct. They have revealed a two-step mechanism in which PvALF potentiates the phaseolin promoter via chromatin modification for subsequent ABA-mediated activation by ABI5, associated with further histone acetylation events (Ng et al, 2006, Ng and Hall, 2008). Over-expression of ABI5 abrogated a requirement for ABA in the PvALF-mediated activation of the phaseolin promoter. The phaseolin promoter could also be activated by FUS3 in an ABA-dependent manner.

#### **4/ End of seed storage protein synthesis**

The end of seed maturation is marked by a down-regulation of maturation phase transcripts, including those encoding transcription factors. There are indications of two mechanisms of down-regulation, the first involving factors that directly repress specific maturation genes, and the second implicating factors that repress upstream (*LEC*) genes.

Firstly, a bZIP factor, named ROM2, peaks in expression after storage protein gene expression in *Phaseolus* (Chern et al. 1996b). ROM2 binds specifically to the phaseolin G-Box in vitro and represses the phaseolin promoter in vivo. An increase in ROM2 DNA binding activity was correlated with the repression of specific maturation genes during seed desiccation (Chern et al. 1996b).

A second mechanism of down-regulation of maturation genes may be via the repression of action of key regulatory proteins such as LEC1 or LEC2. A group of three B3-domain TFs, the *VPI/ABI3-Like* genes (*VAL* genes), has been identified, which may target Sph/RV-containing promoters for repression in the LEC1-B3 transcription network (Suzuki *et al.* 2007). VAL1 is identical to HSI2, independently identified as a transcriptional repressor of sugar-inducible transgenic promoter from sweet potato (Tsukagoshi et al. 2005 and 2007). A SWI/SNF class chromatin remodeling factor, encoded by *PICKLE* (*PKL*), has also been described as a repressor of *LEC1* expression upon germination (Ogas et al. 1999, Rider et al. 2003; Henderson et al. 2004). *pkl* seedlings are capable of expressing embryo-associated traits in non-embryo tissues throughout the plant.

### **INTEGRATION OF NON-GENETIC DETERMINANTS IN THE SEED MATURATION PROGRAMME**

#### **1/ Autonomy of the “embryogenesis regulatory module” due to interconnection between genetic and hormonal controls**

The program of embryogenesis normally accompanies zygote development. However, in many plant species, somatic embryogenesis can be initiated by culturing certain non-zygotic tissues on particular



combinations of plant hormones. Somatic embryos exhibit patterns of cell division and organization identical to those of zygotic embryos, and are also capable of going through the embryo maturation program. Somatic embryogenesis can also be initiated by the ectopic expression of transcription factors such as LEC1 or LEC2. In fact, *lec1* or *lec2* mutants undergo somatic embryogenesis only very inefficiently, suggesting that they are “master regulators” of this process (Gaj et al. 2005). As either LEC1, or LEC2, or certain hormone combinations, suffices to initiate embryogenesis, evidently a degree of redundancy in its regulation has been conserved. This is made possible by the synergistic activation network existing between these major players, (Fig.1), which means that the presence of one component results in the synthesis of the others, and the subsequent unfolding of the program, including modulation of hormone and nutrient concentrations (see Gutierrez et al. (2007) and Santos-Mendoza et al. (2008) for more detailed reviews).

## **2/ Roles played by plant hormones**

Abscisic acid (ABA) is a key hormone during seed maturation (Finkelstein et al. 2002; Nambara and Marion-Poll 2003; Kushiro et al 2003), being required for the acquisition of desiccation tolerance and entry into seed dormancy. Its action is modulated by antagonistic effects of a second hormone, gibberellic acid (GA), on these processes. The central role of ABA signaling is reflected in its interactions with master regulator genes of seed maturation (*ABI3*, *FUS3*, *LEC2* and *LEC1*). For example, ABA levels correlate positively with *FUS3* activity (Gazzarini et al. 2004) and gene activation by *FUS3* and *ABI3* are both strongly enhanced by ABA (Kagaya et al., 2005a, Parcy et al., 2004).

However, seed maturation is not only determined by ABA and combines complex hormonal controls. For example, ectopic *LEC2* expression induces auxin biosynthetic enzymes and auxin-responsive gene activity (Stone et al. 2008), suggesting its action is partly mediated by auxin synthesis. In parallel, its induction causes down-regulation of a cytokinin-responsive promoter (Stone et al. 2008). *LEC1* also acts in concert with auxin to promote embryonic differentiation (Casson et al. 2006). *FUS3* expression is up-regulated by auxin and *FUS3* protein acts as a negative regulator of GA biosynthesis (Gazzarini et al. 2004), both *lec2* and *fus3* mutant embryos possessing increased GA levels (Curaba et al. 2004).

## **3/ Role-played by metabolic regulation**

Another factor determining the course of seed maturation is the concentration of certain metabolites. Among these metabolites, sugar and specifically the sucrose/hexose ratio provides signals for the onset of maturation phase (Weber et al. 2005). Sucrose signals control storage and differentiation processes through the regulation of gene expression and activity (Gibson 2005; Rolland et al. 2006). For example, Ohto et al. (2005) have shown that the *apetala2* mutants, affected in seed mass and characterized by an increase of seed cell size and number, have an increased hexose/sucrose ratio that results in an extension of the period of cell division.



Casson et al. (2006) have also demonstrated that LEC1 promotes embryonic differentiation in part via effects on sugar levels.

Other metabolic signals play important roles in seed maturation: ATP and oxygen levels are correlated with storage-related processes (Borisjuk et al. 2005; Rolletschek et al. 2005a; Weber et al. 2005 and Vigeolas et al. 2003). Nitrogen supplied by the phloem is limiting for storage protein synthesis (Rolletschek et al. 2005b) possibly through the control of asparagine content in the developing cotyledons (Hernandez-Sebastia et al. 2005).

## **POTENTIAL OF GLOBAL POST-GENOMICS APPROACHES**

### **1. Comparative genomics**

The observation of a high diversity of seed storage compounds in dicotyledons raises the question as to whether a wide range of regulatory proteins also exists. Recently, the framework of regulation established essentially using *A. thaliana*, has been compared with genome wide expression analyses performed on various other model plants. These approaches highlight new unknown mechanisms and allow a distinction between conserved and non-conserved seed maturation processes. Gutierrez et al. (2006) have identified 13 new gene expression regulators specifically expressed during seed maturation in flax using a cDNA-AFLP approach and confirmed by qRT-PCR analysis. Out of these regulators, the authors focused on two, a non LEC1-type gene and a MYB factor. A putative orthologue to the non LEC1-type has now been identified in *A. thaliana* seed maturation and could be part of a ubiquitous mechanism (Gutierrez et al. 2006). In contrast, the MYB factor appears to represent a non-ubiquitous regulatory mechanism more specific to the seed maturation mechanism of flax.

Studies on grain development of the model legume *M. truncatula* were carried out using a *M. truncatula* gene expression atlas, based on a 50,900 sequence Affymetrix array which provides a global view of gene expression in all major organ systems including seeds (Benedito et al., 2008). This revealed significant divergence in developmental expression profiles between putatively orthologous TF genes from *M. truncatula* and *A. thaliana*. These data also represent a resource for identifying *M. truncatula* seed-specific genes. In another study in this species, a profiling of TF gene expression, using quantitative RT-PCR has recently allowed the identification of more 160 TFs implicated in different phases during seed development (Verdier et al. 2008). The TF profiles were subjected to hierarchical clustering along with expression data for major *M. truncatula* seed storage proteins. The six different clusters of TF and putative target genes obtained correspond to expression profiles at different developmental stages, from late embryogenesis to the desiccation phase. The maturation phase was represented by three different clusters corresponding to the sequential accumulation of the storage proteins vicilin (TC100302) (16-20 DAP), legumin K (TC100253) (20-24 DAP) and legumin A (TC100252) (24-36 DAP). This phasing in accumulation of different storage protein families, observed generally amongst legumes, contrasts with the synchronous accumulation of the same protein types in *A. thaliana*. These



three clusters identified include possible legume-specific orthologues of regulators of seed storage protein synthesis including B3 domain, bZIP, DOF and MYB family members (Verdier et al. 2008).

## 2. Combining genomics approaches

The combination of different genomics approaches is providing new insights into the regulation of seed maturation. In *A. thaliana*, Ruuska et al (2002) used cDNA microarrays to compare wild-type and mutant *wrinkled1* (*wri1*) seeds that have an 80% reduction in oil content. mRNA expression profiles could be classified into a few discrete patterns corresponding to specific stages in seed development. Despite major changes in metabolism between *wri1* and wild-type seeds (Focks and Benning, 1998), <1% of seed-expressed genes differed by more than two-fold, and most of these were involved in central lipid and carbohydrate metabolism, supporting a role for WRI1, a putative transcription factor, in metabolic regulation.

Using a combination of proteomics and transcriptomics analyses of developing *Medicago truncatula* seeds, a compartmentalization of amino acid metabolic pathways within distinct seed tissues was documented that is thought to optimize amino acid availability for efficient storage protein synthesis (Gallardo et al., 2007).

By combining metabolomics with transcriptomics data, and using unsupervised clustering analyses, novel examples of the dynamics and specializations of metabolic pathways during seed development are coming to light. Metabolic and transcriptomic profiling of *Pisum sativum* seeds lacking a major albumin, PA2, (Vigeolas et al., 2008), revealed an unsuspected role for this protein in regulating polyamine metabolism, important in development and stress responses.

Further integrating post-genomics studies on different model plants will enable us to distinguish between conserved and non-conserved processes and to better understand the complex interplay of events regulating seed reserve synthesis.

## REFERENCES

- Alemanno, L., Devic, M., Niemenak, N., Sanier, C., Guilleminot, J., Rio, M., Verdeil, J.L., Montoro, P. (2008) Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. *Planta* 227(4):853-66.
- Bassuner, R., Manteuffel, R., Muntz, K., Puchel, M., Schmidt, P. and Weber, E. (1983) Analysis of in vivo and in vitro globulin formation during cotyledon development of field beans (*Vicia faba* L. var. *minor*). *Biochemie und Physiologie der Pflanzen* 178(8): 665-684.
- Bäumlein, H., Boerjan, W., Nagy, I., Bassfüner, R., Van Montagu, M., Inzé, D. and Wobus, U. (1991) A novel seed protein gene from *Vicia faba* is developmentally regulated in transgenic tobacco and *Arabidopsis* plants. *Molecular and General Genetics* 225(3): 459-467.





- Bäumlein, H., Misera, S., Luerksen, H., Kolle, K., Horstmann, C., Wobus, U., Müller, A. (1994) The FUS3 gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. *Plant J.* 6: 367-387.
- Bäumlein, H., Nagy, I., Villarroel, R., Inzé, D., Wobus, U. (1992) Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* 2(2):233-9.
- Bäumlein, H., Wobus, U., Pustell, J., Kafatos, F.C. (1986) The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucleic Acids Res.* 14(6):2707-20.
- Beach, L.R., Spencer, D., Randall, P.J., Higgins, T.J. (1985) Transcriptional and post-transcriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. *Nucleic Acids Res.* 13(3):999-1013.
- Benedito, V.A., Torres-Jerez, I., Murray, J.D., Andriankaja, A., Allen, S. et al (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* June 4, PMID: 18410479 (in press).
- Bensmihen, S., Rippa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F., Giraudat, J., Parcy, F. (2002) The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell* 14(6):1391-403.
- Bobb, A.J., Chern, M.S., Bustos, M.M. (1997) Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. *Nucleic Acids Res.* 25(3):641-7.
- Bobb, A.J., Eiben, H.G., Bustos, M.M. (1995) PvAlf, an embryo-specific acidic transcriptional activator enhances gene expression from phaseolin and phytohemagglutinin promoters. *Plant J.* 8(3):331-43.
- Borisjuk, L., Nguyen, T.H., Neuberger, T., Rutten, T., Tschiersch, H. et al (2005) Gradients of lipid storage, photosynthesis and plastid differentiation in developing soybean seeds. *New Phytol.* 167(3):761-76.
- Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., Harada, J.J. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci U S A.* 103(9):3468-73.
- Britten, R.J. and Davidson, E.H. (1969) Gene Regulation for Higher Cells: A Theory. *Science* 1969: 349-357.
- Busk, P.K. and Pagès, M. (1998) Regulation of abscisic acid-induced transcription. *Plant Mol. Biol.* 37(3): 425-435.
- Casson, S.A. and Lindsey, K. (2006) The turnip mutant of *Arabidopsis* reveals that LEAFY COTYLEDON1 expression mediates the effects of auxin and sugars to promote embryonic cell identity. *Plant Physiol.* 142(2):526-41.
- Castle, L.A., Meinke, D.W. (1994) A FUSCA gene of *Arabidopsis* encodes a novel protein essential for plant development. *Plant Cell* 6(1):25-41.



- Cernac, A. and Benning, C. (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *Plant J.* 40(4):575-85.
- Chandrasekharan MB, Bishop KJ, Hall TC (2003) Module-specific regulation of the beta-phaseolin promoter during embryogenesis. *Plant Journal* 33: 853-866
- Chern, M.S., Bobb, A.J., Bustos, M.M. (1996b) The regulator of MAT2 (ROM2) protein binds to early maturation promoters and represses PvALF-activated transcription. *Plant Cell.* 8(2):305-21.
- Chern, M.S., Eiben, H.G., Bustos, M.M. (1996a) The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. *Plant J.* 10(1):135-48.
- Curaba, J., Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M., Vachon, G. (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. *Plant Physiol.* 136(3):3660-9.
- De Pater, S., Pham, K., Chua, N.H., Memelink, J., Kijne, J. (1993) A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. *Plant Cell.* 5(8):877-86.
- Edwards, D., Murray, J.A., Smith, A.G. (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in Arabidopsis. *Plant Physiol.* 117(3):1015-22.
- Evans, I.M., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) Regulation of the transcription of storage-protein mRNA in nuclei isolated from developing pea (*Pisum sativum* L.) cotyledons. *Planta* 160(6): 559-568.
- Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K. and Rask, L. (1999) Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol Biol.* 40(4):699-709.
- Ezcurra, I., Wycliffe, P., Nehlin, L., Ellerström, M., Rask, L. (2000) Transactivation of the Brassica napus napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. *Plant J.* 24(1):57-66.
- Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiology* 142: 839-854
- Finkelstein, R.R. and Gibson, S.I. (2002) ABA and sugar interactions regulating development: cross-talk or voices in a crowd. *Curr Opin Plant Biol.* 5(1):26-32.
- Focks N., Benning C. (1998) *Wrinkled1*: A novel low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate. *Plant Physiol.* 118: 91-101.



- Gaj, M.D., Zhang, S., Harada, J.J. and Lemaux, P.G. (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. *Planta* 222(6):977-88.
- Garcia, D., Gerald, J.N.F. and Berger, F. (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* 17(1):52-60.
- Gatehouse, J.A., Evans, I.M., Bown, D., Croy, R.R., Boulter, D. (1982) Control of storage-protein synthesis during seed development in pea (*Pisum sativum* L.). *Biochem J.* 208(1):119-27.
- Gazzarrini, S., Tsuchiya, Y., Luma, S., Okamoto, M., McCourt, P. (2004) The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones Giberellin and Abscisic Acid. *Dev. Cell* 7: 373-385.
- Gibson, S.I. (2005) Control of plant development and gene expression by sugar signaling. *Curr Opin Plant Biol.* 8(1):93-102.
- Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parey, F., Goodman, H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4(10):1251-61.
- Greenwood, J.S. and Chrispeels, M.J. (1985) Correct Targeting of the Bean Storage Protein Phaseolin in the Seeds of Transformed Tobacco. *Plant Physiol.* 79(1): 65–71.
- Gutierrez, L., Conejero, G., Castelain, M., Guénin, S., Verdeil, J.L., Thomasset, B., Van Wuytswinkel, O. (2006) Identification of new gene expression regulators specifically expressed during plant seed maturation. *J Exp Bot.* 57(9):1919-32.
- Gutierrez, L., Van Wuytswinkel, O., Castelain, M., Bellini, C. (2007) Combined networks regulating seed maturation. *Trends Plant Sci.* 12(7):294-300.
- Harada, J.J., Barker, S.J. and Goldberg, R.B. (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell.* 1(4):415-25.
- Hardtke, C.S. and Berleth, T. (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J.* 17(5):1405-11.
- Henderson, J.T., Li, H.C., Rider, S.D., Mordhorst, A.P., Romero-Severson, J., Cheng, J.C., Robey, J., Sung, Z.R., De Vries, S.C., Ogas, J. (2004) PICKLE acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. *Plant Physiol.* 134(3):995-1005.
- Hernández-Sebastià, C., Marsolais, F., Saravitz, C., Israel, D., Dewey, R.E., Huber, S.C. (2005) Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. *J Exp Bot.* 56(417):1951-63.



- Hobo, T., Kowyama, Y., Hattori, T. (1999) A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. *Proc Natl Acad Sci U S A.* 96(26):15348-53.
- Hoefgen, R; Nikiforova, VJ (2008) Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiologia Plantarum*, 132 (2): 190-198
- Huntley, R.P. and Murray, J.A.H. (1999) The plant cell cycle. *Current Opinion in Plant Biology* 2(6): 440-446.
- Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., Yamamoto, A., Hattori, T. (2005a) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in *Arabidopsis*. *Plant Cell Physiol.* 46(2):300-11.
- Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T. (2005b) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46(3):399-406.
- Kawagoe, Y. and Mura, N. (1996) A novel basic region/helix-loop-helix protein binds to a G-box motif CACGTG of the bean seed storage protein  $\beta$ -phaseolin gene. *Plant Science* 116(1): 47-57.
- Kawagoe, Y., Campbell, B.R., Murai, N. (1994) Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein beta-phaseolin gene. *Plant J.* 5(6):885-90.
- Keith, K., Kraml, M., Dengler, N.G., McCourt, P.(1994) *fusca3*: A Heterochronic Mutation Affecting Late Embryo Development in *Arabidopsis*. *Plant Cell* 6(5):589-600.
- Kroj, T., Savino, G., Valon, C., Giraudat, J., Parcy, F. (2003) Regulation of storage protein gene expression in *Arabidopsis*. *Development.* 130(24):6065-73.
- Kushiro, T., Nambara, E. and McCourt P. (2003) Hormone evolution: The key to signalling. *Nature* 422(6928):122.
- Kwong, R.W., Bui, A.Q., Lee, H., Kwong, L.W., Fischer, R.L., Goldberg, R.B., Harada, J.J. (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15(1):5-18.
- Ladin, B.F., Tierney, M.L., Meinke, D.W., Hosángadi, P., Veith, M., and Beachy, R.N. (1987) Developmental Regulation of  $\beta$ -Conglycinin in Soybean Axes and Cotyledons. *Plant Physiol.* 84(1): 35–41.
- Lara, P., Oñate-Sánchez, L., Abraham, Z., Ferrándiz, C., Díaz, I., Carbonero, P., Vicente-Carbajosa, J. (2003) Synergistic activation of seed storage protein gene expression in *Arabidopsis* by ABI3 and two bZIPs related to OPAQUE2. *J Biol Chem.* 278(23):21003-11.
- Li, G., Chandrasekharan, M.B., Wolffe, A.P., Hall, T.C. (2001) Chromatin structure and phaseolin gene regulation. *Plant Mol Biol.* 46(2):121-9.





- Lotan, T., Ohto, M., Yee, K., West, M., Lo, R., Kwong, R., Yamagishi, K., Fischer, R., Goldberg, R., Harada, J.J., (1998) *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93: 1195-1205.
- Maity, S.N. and De Crombrugghe, B. (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci.* 23(5):174-8.
- 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(5):895-905.
- McCourt, P. (1999) Analysis of hormone signalling. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 219-243.
- Meinke, D.W. (1992) A Homoeotic Mutant of *Arabidopsis thaliana* with Leafy Cotyledons. *Science* 258(5088):1647-1650.
- Meinke, D.W., Chen, J. and Beachy, R.N. (1981) Expression of storage-protein genes during soybean seed development. *Planta* 153(2): 130-139.
- Meinke, D.W., Franzmann, L.H., Nickle, T.C., Yeung, E.C. (1994) Leafy Cotyledon Mutants of *Arabidopsis*. *Plant Cell* 6(8):1049-1064.
- Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.P., Bäumlein, H., Conrad, U. (2004) Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta* 219(1):158-66.
- Mounet, F; Lemaire-Chamley, M; Maucourt, M; Cabasson, C; Giraudel, JL et al. (2007) Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics*, 3 (3): 273-288.
- Nakamura, S., Lynch, T.J., Finkelstein, R.R. (2001) Physical interactions between ABA response loci of *Arabidopsis*. *Plant J.* 2001 26(6): 627-35.
- Nambara, E. and Marion-Poll, A. (2003) ABA action and interactions in seeds. *Trends Plant Sci.* 8(5):213-7.
- Ng, D.W. and Hall, T.C. (2008) PvALF and FUS3 activate expression from the phaseolin promoter by different mechanisms. *Plant Mol Biol.* 66(3):233-44.
- Ng, D.W., Chandrasekharan, M.B., Hall, T.C. (2006) Ordered histone modifications are associated with transcriptional poising and activation of the phaseolin promoter. *Plant Cell* 18(1):119-32.
- Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallon, B.J., Fischer, R.L., Sims, T.L., Drews, G.N. and Goldberg, R.B. (1989) Characterization of the glycinin gene family in soybean. *Plant Cell.* 1(3):313-28.



- Ogas, J., Kaufmann, S., Henderson, J., Somerville, C. (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. *Proc Natl Acad Sci U S A*. 96(24):13839-44.
- Ohto, M.A., Fischer, R.L., Goldberg, R.B., Nakamura, K., Harada, J.J. (2005) Control of seed mass by APETALA2. *Proc Natl Acad Sci U S A*. 102(8):3123-8.
- Parcy, F., Valon, C., Kohara, A., Miséra, S., Giraudat, J. (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell*. 9(8):1265-77.
- Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., Giraudat, J. (1994) Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell*. 6(11):1567-82.
- Pang, P.P., Pruitt, R.E. and Meyerowitz, E.M. (1988) Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana. *Plant Mol. Biol.* 11(6): 805-820.
- Pla, M., Vilardell, J., Guiltinan, M.J., Marcotte, W.R., Niogret, M.F., Quatrano, R.S., Pagès M. (1993) The cis-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene rab28. *Plant Mol Biol*. 21(2):259-66.
- Qu, L.J. and Zhu, Y.X. (2006) Transcription factor families in Arabidopsis: major progress and outstanding issues for future research. *Current Opinion in Plant Biology* 9(5): 544-549.
- Reidt, W., Ellerström, M., Kölle, K., Tewes, A., Tiedemann, J., Altschmied, L. and Bäumlein, H. (2001) FUS3-dependent gene regulation during late embryogenesis. *Journal of Plant Physiology* 158(4): 411-418.
- Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L., Bäumlein, H. (2000) Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J*. 21(5):401-8.
- Rider, S.D., 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. *Plant J*. 35(1):33-43.
- Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology* 3(5): 423-434.
- Rolland, F., Baena-Gonzalez, E. and Sheen, J. (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol*. 57:675-709.



- Rolletschek, H., Koch, K., Wobus, U. and Borisjuk, L. (2005a) Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. *Plant J.* 42(1):69-83.
- Rolletschek, H., Hosein, F., Miranda, M., Heim, U., Götz, K.P., Schlereth, A., Borisjuk, L., Saalbach, I., Wobus, U. and Weber, H. (2005b) Ectopic expression of an amino acid transporter (VfAAP1) in seeds of *Vicia narbonensis* and pea increases storage proteins. *Plant Physiol.* 137(4):1236-49.
- Ruuska, S.A., Girke, T., Benning, C., Ohlrogge, J.B. (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* 14(6):1191-206.
- Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M. and Lepiniec, L. (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Lett.* 579(21):4666-70.
- Santos-Mendoza M, Dubreucq B, Baud S, Parcy F., Caboche M., Lepiniec L. (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J.* 54: 608-620
- Shotwell, M.A., and Larkins, B.A. (1989) The molecular biology and biochemistry of seed storage proteins. In *The Biochemistry of Plants*, Vol. 15, A. Marcus, ed. (San Diego, CA: Academic Press), pp. 297–345.
- Singh, K.B., Foley, R.C. and Oñate-Sánchez, L. (2002) Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology* 5(5): 430-436.
- Stålberg, K., Ellerstöm, M., Ezcurra, I., Ablov, S., Rask, L. (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. *Planta* 199(4):515-9.
- Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.F., Fischer, R.L., Goldberg, R.B., Harada, J.J. (2008) *Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci U S A.* 105(8):3151-6.
- Suzuki, M., Wang, H.H., McCarty, D.R. (2007) Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol.* 143(2):902-11.
- Sun, T. and Gubler, F. (2004) Molecular mechanism of gibberellin signalling in plants. *Annual Review of Plant Biology* 55: 197-223.
- To, A., Valon, C., Savino, G., Guillemot, J., Devic, M., Giraudat, J., Parcy, F. (2006) A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. *Plant Cell* 18(7):1642-51.
- Tsuchiya, Y., Nambara, E., Naito, S., McCourt, P. (2004) The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in *Arabidopsis*. *Plant J.* 37(1):73-81.



- 
- Tsukagoshi, H., Morikami, A., Nakamura, K. (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in *Arabidopsis* seedlings. *Proc Natl Acad Sci U S A.* 104(7):2543-7.
- Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A., Nakamura, K. (2005) Analysis of a sugar response mutant of *Arabidopsis* identified a novel B3 domain protein that functions as an active transcriptional repressor. *Plant Physiol.* 138(2):675-85.
- Verdier, J., Kakar, K., Gallardo, K., Le Signor, C., Aubert, G., Schlereth, A., Town, C.D., Udvardi, M.K., Thompson, R.D. (2008) Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol Biol.* 67(6):567-580.
- Vicente-Carbajosa, J., Carbonero P. (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int. J. Dev. Biol.* 49: 645-651 (2005)
- Vigeolas, H; Chinoy, C; Zuther, E; Blessington, B; Geigenberger, P; Domoney, C (2008) Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds. *Plant Physiol.* 146(1):74-82.
- Vigeolas, H., Van Dongen, J.T., Waldeck, P., Huhn, D. and Geigenberger, P. (2003) Lipid storage metabolism is limited by the prevailing low oxygen concentrations within developing seeds of oilseed rape. *Plant Physiol.* 133(4):2048-60.
- Walling, L., Drews, G.N. and Goldberg, R.B. (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci U S A.* 83(7): 2123–2127.
- Weber, H., Borisjuk, L. and Wobus, U. (2005) Molecular physiology of legume seed development. *Annu Rev Plant Biol.* 56:253-79.
- West, M., Yee K., Danao J., Zimmerman J. Fischer R., Goldberg R., Harada J.J. (1994) LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* 6: 1731-1745.





## 5. OBJECTIFS ET DÉMARCHE DE LA THÈSE

### 5.1 Objectifs de la thèse

L'objectif général de ce travail de thèse est de **mieux comprendre les mécanismes de la régulation transcriptionnelle de type gène spécifique**, donc faisant intervenir des facteurs de transcription durant le développement de la graine des légumineuses et plus particulièrement **durant la phase de remplissage**.

Les données bibliographiques disponibles montrent que les protéines de réserve sont régulées de manière transcriptionnelle chez les légumineuses mais peu d'études ont tenté d'identifier les facteurs de transcription impliqués dans cette phase critique pour la qualité finale du grain. Ces études étaient focalisées sur l'espèce modèle *A. thaliana* où quelques gènes ont été identifiés comme participant à cette régulation. Cependant, *A. thaliana* est phylogénétiquement éloignée des légumineuses et même si le cadre de la régulation est probablement similaire entre ces espèces, plusieurs différences sont observables. Par exemple, la synthèse et l'accumulation des protéines de réserve chez *A. thaliana* est synchrone, ce qui constitue la différence majeure avec les légumineuses qui présentent un décalage dans l'expression et l'accumulation des différentes classes de protéines de réserve, décalage probablement régulé transcriptionnellement.

L'intérêt scientifique de ces travaux se situe à deux niveaux. D'abord **d'un point de vue fondamental**, il s'agit d'améliorer les connaissances sur les mécanismes transcriptionnels du développement de la graine chez les légumineuses. L'amélioration de ces connaissances passe par l'identification des mécanismes communs entre les légumineuses et ceux déjà connus chez *A. thaliana*, et par l'identification des mécanismes, et donc les gènes, qui diffèrent entre ces espèces. Ensuite **d'un point de vue appliqué**, il s'agit d'identifier des gènes directement impliqués dans des processus affectant des caractères agronomiques de qualité finale du grain tels que le poids du grain ou sa teneur en protéines.

Ce projet de recherche a été initié au sein de l'UMR Génétique et Ecophysiologie des Légumineuses (INRA UMR-LEG, Dijon) et s'inscrit dans l'axe de recherche "Morphogenèse des graines et accumulation des réserves". Ce travail a été réalisé dans l'équipe "Développement et qualité de la graine" de l'UMR-LEG et cofinancé par le département INRA de Génétique et Amélioration des Plantes et par le projet européen FP6, Grain Legumes Integrated Project (<http://eugrainlegumes.org>). Ce travail s'insère dans deux modules de ce programme de recherche conduit à l'échelle européenne : les modules WP5.2 "Mutagenesis and reverse genetics" et WP3.2 "Novel approaches to alter seed composition".

### 5.2 Démarche de ce travail de thèse

Les résultats issus de ces travaux peuvent être résumés en trois parties, présentées ci-dessous. Au cours de cette étude, les résultats des trois premiers chapitres ont été obtenus en utilisant la légumineuse modèle *Medicago truncatula* cv. Jemalong, lignée A17. La lignée R108, également du cv. Jemalong, a été utilisée pour les expériences de transgénèse. En revanche, le chapitre 4 comporte des résultats qui ont été obtenus chez l'espèce



d'intérêt agronomique, *Pisum sativum* cv. Caméor considéré comme lignée de référence pour les travaux de génétique et de génomique sur le pois.

La première partie de ce travail (Chapitre 1) a consisté à **élaborer un cadre descriptif des mécanismes moléculaires durant le développement de la graine**. Ce travail s'appuie sur les données obtenues en transcriptomique sur le suivi d'expression de 51k gènes au cours du développement de la graine de *M. truncatula*.

Puis, la deuxième partie (Chapitre 2) a consisté à **identifier des facteurs de transcription impliqués dans le développement de la graine des légumineuses et pouvant réguler la phase de remplissage**, et en particulier la synthèse des protéines de réserve. Cette étude a été réalisée grâce à une approche de suivi d'expression de plus de 700 facteurs de transcription pendant le développement de la graine.

Enfin, la troisième partie a consisté en **l'analyse fonctionnelle d'un facteur de transcription précédemment identifié, un gène *LECI-LIKE***. Dans un premier temps, **son rôle dans le développement de la graine de *M. truncatula*** a été analysé (Chapitre 3). Dans un second temps, nous avons étudié **l'impact de ce gène sur la qualité de la graine de pois** (Chapitre 4).



**CHAPITRE 1 : Exploration des  
gènes de *M. truncatula*  
exprimés lors du  
développement de la graine**



## CHAPITRE 1

# Exploration des gènes de *M. truncatula* exprimés lors du développement de la graine

### 1. OBJECTIFS ET DEMARCHE

La génomique fonctionnelle est l'un des enjeux de la biologie actuelle et a pour but de déterminer à grande échelle les mécanismes biologiques impliqués dans les différents processus développementaux. Les avancées récentes dans les approches globales telles que la transcriptomique ou la protéomique permettent aujourd'hui d'identifier les mécanismes moléculaires simultanément mis en jeu dans ces processus. De plus en plus de processus biologiques sont étudiés grâce à cette approche : le contrôle du développement hormonal (Che *et al.* 2002 et Goda *et al.* 2002), les interactions plante-pathogène (Caldo *et al.* 2004), la réponse au stress environnemental (Rossel *et al.* 2002, Cheong *et al.* 2002 et Cooper *et al.* 2003), le gravitropisme (Moseyko *et al.* 2002), le cycle cellulaire (Menges *et al.* 2002), les rythmes circadiens (Harmer *et al.* 2000). Des processus développementaux sont également disséqués tels que le développement floral (Schmid *et al.* 2003 et Wellmer *et al.* 2004), le développement de la graine (Hunter *et al.* 2002, Ruuska *et al.* 2002, Zhu *et al.* 2003, Buitink *et al.* 2006, Gallardo *et al.* 2007) ou le développement des cellules méristématiques racinaires (Holmes *et al.* 2008). Le développement des outils de traitement informatique permet aujourd'hui de réaliser un suivi de l'expression de gènes dans différents tissus et également au cours de différents stades de développement, en répondant toujours aux mêmes conditions expérimentales. Ces analyses de plus en plus exhaustives sont regroupées sous le terme d'atlas de gènes ("gene atlas") : elles servent à cataloguer les niveaux d'expression des gènes dans les différents tissus de la plante et durant différents processus biologiques. De nombreux atlas de gènes ont vu le jour comme par exemple chez le maïs (Cho *et al.* 2002), le riz (Zhu *et al.* 2003) ou l'orge (Druka *et al.* 2006). En plus de générer une base de données descriptive sur l'expression et la régulation des gènes intra-espèce, ces atlas sont également une source d'informations pour étudier la conservation des mécanismes inter-espèces.

Chez les dicotylédones, le seul atlas de gènes disponible, à ce jour, est celui d'*A. thaliana* (Czechowski *et al.* 2005 et Schmid *et al.* 2005). Cependant, le développement de la puce à ADN Affymetrix de *M. truncatula* contenant approximativement 51000 (51k) gènes a permis la naissance d'un projet collaboratif ayant pour but de générer un atlas de gènes pour cette légumineuse modèle. Ce projet a été centralisé par Mickael Udvardi (Noble Foundation, Ardmore, USA). Plusieurs partenaires ont été sollicités pour participer à ce projet en fonction de leur thématique de recherche. Ainsi, nous avons contribué à ce projet via l'étude du développement de la graine.





Ce chapitre présente dans une première partie la description globale de l'atlas de gènes de *M. truncatula*, sous la forme d'une publication. La seconde partie analyse plus spécifiquement le développement de la graine et les différents mécanismes moléculaires sous jacents à ce processus biologique.



## **2. UN ATLAS DE L'EXPRESSION DES GENES CHEZ LA LEGUMINEUSE MODELE *M. TRUNCATULA***

Ce travail se réfère à un article publié dans la revue *The Plant Journal* intitulé :

### **A gene expression atlas of the model legume *Medicago truncatula***

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, **Verdier J**, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK

*The Plant Journal*, June 2008, PMID: 18410479

#### Résumé :

Les légumineuses ont joué un rôle central dans le développement des civilisations et représentent aujourd'hui approximativement un tiers de la production agricole mondiale. Malheureusement, pour de nombreuses raisons, les légumineuses cultivées ne constituent pas de bons modèles pour la recherche génomique. Par conséquent, la plante *M. truncatula* a été adoptée comme espèce modèle, en raison d'un génome diploïde de petite taille. Pour développer nos connaissances sur cette plante modèle, nous avons généré un atlas de l'expression des gènes qui fournit une vue globale du transcriptome de *M. truncatula* dans tous les organes majeurs, en mettant l'accent sur le développement des nodules et des graines. Cet atlas révèle de grandes différences dans l'expression des gènes entre les différents organes. Ces différences s'accompagnent de changement dans l'expression de gènes régulateurs clefs comme les facteurs de transcription, qui orchestrent vraisemblablement les programmes génétiques durant le développement et la différenciation des organes. De manière intéressante, beaucoup de gènes spécifiques des légumineuses sont préférentiellement exprimés dans les nodules fixateurs d'azote, ce qui souligne la spécificité de cet organe. Une analyse comparative des transcriptomes de *M. truncatula* et *A. thaliana* a révélé d'importantes variations dans le profil d'expression de gènes orthologues associés aux aspects développementaux. Ceci indique que les analyses phylogénétiques seules ne sont pas suffisantes pour prédire la fonction des gènes entre différentes espèces. Les données, décrites ici, représentent une ressource sans précédent pour développer et accélérer les découvertes en génomique fonctionnelle des légumineuses.



## TECHNICAL ADVANCE

# A gene expression atlas of the model legume *Medicago truncatula*

Vagner A. Benedito<sup>1</sup>, Ivone Torres-Jerez<sup>1</sup>, Jeremy D. Murray<sup>1</sup>, Andry Andriankaja<sup>1</sup>, Stacy Allen<sup>1</sup>, Klementina Kakar<sup>2</sup>, Maren Wandrey<sup>2</sup>, Jérôme Verdier<sup>3</sup>, Hélène Zuber<sup>3</sup>, Thomas Ott<sup>4</sup>, Sandra Moreau<sup>4</sup>, Andreas Niebel<sup>4</sup>, Tancred Frickey<sup>5</sup>, Georg Weiller<sup>5</sup>, Ji He<sup>1</sup>, Xinbin Dai<sup>1</sup>, Patrick X. Zhao<sup>1</sup>, Yuhong Tang<sup>1</sup> and Michael K. Udvardi<sup>1,\*</sup>

<sup>1</sup>Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA,

<sup>2</sup>Max-Planck Institute of Molecular Plant Physiology, Am Muhlenberg 1, 14476 Golm, Germany,

<sup>3</sup>INRA-URLEG, Unite de Recherche sur les Legumineuses, BP 86510, F-21065 Dijon Cedex, France,

<sup>4</sup>INRA-CNRS, 31326 Castanet-Tolosan, France, and

<sup>5</sup>Research School of Biological Sciences, The Australian National University, GPO Box 475, Canberra, ACT 2601, Australia

Received 12 February 2008; accepted 13 March 2008.

\*For correspondence (fax +1 580 224 6692; e-mail mudvardi@noble.org).

## Summary

Legumes played central roles in the development of agriculture and civilization, and today account for approximately one-third of the world's primary crop production. Unfortunately, most cultivated legumes are poor model systems for genomic research. Therefore, *Medicago truncatula*, which has a relatively small diploid genome, has been adopted as a model species for legume genomics. To enhance its value as a model, we have generated a gene expression atlas that provides a global view of gene expression in all major organ systems of this species, with special emphasis on nodule and seed development. The atlas reveals massive differences in gene expression between organs that are accompanied by changes in the expression of key regulatory genes, such as transcription factor genes, which presumably orchestrate genetic reprogramming during development and differentiation. Interestingly, many legume-specific genes are preferentially expressed in nitrogen-fixing nodules, indicating that evolution endowed them with special roles in this unique and important organ. Comparative transcriptome analysis of *Medicago* versus *Arabidopsis* revealed significant divergence in developmental expression profiles of orthologous genes, which indicates that phylogenetic analysis alone is insufficient to predict the function of orthologs in different species. The data presented here represent an unparalleled resource for legume functional genomics, which will accelerate discoveries in legume biology.

**Keywords:** *Medicago truncatula*, transcriptome, development, nodule, seed.

## Introduction

Legumes (family Fabaceae) are second only to grasses (Gramineae) in importance to humans as a source of food, feed for livestock, and raw materials for industry (Graham and Vance, 2003). Legumes account for one-third of the world's primary crop production and are key to sustainable agriculture because they can 'fix' nitrogen (reduce N<sub>2</sub> to NH<sub>3</sub>) in a symbiotic association with bacteria called rhizobia, providing crops with a free and renewable source of nitrogen. It is estimated that 40–60 million tonnes of N are fixed annually by cultivated legumes (Smil, 1999), saving about

US\$10 billion on nitrogen fertilizer (Graham and Vance, 2003).

Symbiotic nitrogen fixation (SNF) in legumes takes place in specialized organs called nodules which develop from root cortical cells that start dividing following signal exchanges between the plant roots and rhizobia in the soil (Brewin, 1991; Long, 2001; Oldroyd and Downie, 2004). Rhizobia gain entry into cortical cells of developing nodules via an infection thread, which traverses the epidermal root hair cell that makes first contact with the bacteria, and



subsequently ramifies throughout the cortical tissue. Concomitantly, dividing cortical cells form a nodule primordium inside which a new meristem is initiated that drives nodule organogenesis. Bacteria are then released from infection threads into the cytoplasm of cortical cells via endocytosis, which leaves the bacteria surrounded by a host membrane called the symbiosome membrane (Udvardi and Day, 1997). Within the resulting organelle, called the symbiosome, the bacteria multiply and ultimately differentiate into their nitrogen-fixing 'bacteroid' state. Infected cortical cells typically contain thousands of symbiosomes, each containing one or a few bacteroids. Transcriptome analyses have identified hundreds of plant and bacterial genes that are differentially expressed during nodule development and differentiation (Becker *et al.*, 2004; Colebatch *et al.*, 2002, 2004; El-Yahyaoui *et al.*, 2004; Mitra *et al.*, 2004; Uchiyama *et al.*, 2004), although genome-wide studies of plant gene expression during nodulation are yet to be reported. Likewise, although transcriptome studies have been carried out on different organs in multiple legume species under a variety of experimental conditions (Colebatch *et al.*, 2004; El-Yahyaoui *et al.*, 2004; Zabala *et al.*, 2006), no single study has yet brought together genome-wide data for all major plant organs of a single species.

## Results and discussion

To create a comprehensive gene expression atlas for *Medicago truncatula*, we utilized the new Affymetrix GeneChip Medicago Genome Array, which contains 50 900 probe sets representing the majority of genes in this species. Gene expression values were obtained from three independent biological replicates of each of the major organ systems: roots, nodules, stems, petioles, leaf blades, vegetative buds, flowers, and seed pods. In addition, multiple stages of nodule and seed development were profiled to obtain greater insight into the transcriptional programs that underlie the development of these two organs, which are the foci of most legume research. The fraction of genes for which transcripts were detected (called 'present' by dCHIP; Li and Wong, 2001) was similar in all samples, and ranged from 55.2% in seeds to 63.3% in roots (Figure S1). Similar results were obtained recently for *Arabidopsis* (Schmid *et al.*, 2005). However, *Arabidopsis* cannot nodulate and provides no information about symbiotic nitrogen fixation.

No transcripts were detected for 13.9% of putative genes in any of the organs tested. The majority of these correspond to genes annotated from genomic sequences by the International Medicago Genome Annotation Group (IMGAG; Town, 2006). No evidence for expression of 23.3% of IMGAG-annotated genes was found, corresponding to 4370 out of 18731 probe sets. It is possible that some of these have been incorrectly annotated as genes, although more thorough sampling of the transcriptome, by encom-

passing a wider range of developmental stages, growth, and stress conditions, together with more sensitive measurement devices such as quantitative (q)RT-PCR, will be required to conclude that such DNA is not transcribed under any conditions. In contrast to the results obtained from probe sets designed from IMGAG-annotated genes, only 8.4% of probe sets derived from expressed sequence tag (EST)/cDNA sequences did not detect gene transcripts in any of the organs tested, which presumably reflects the inherent bias in EST data towards more highly expressed genes.

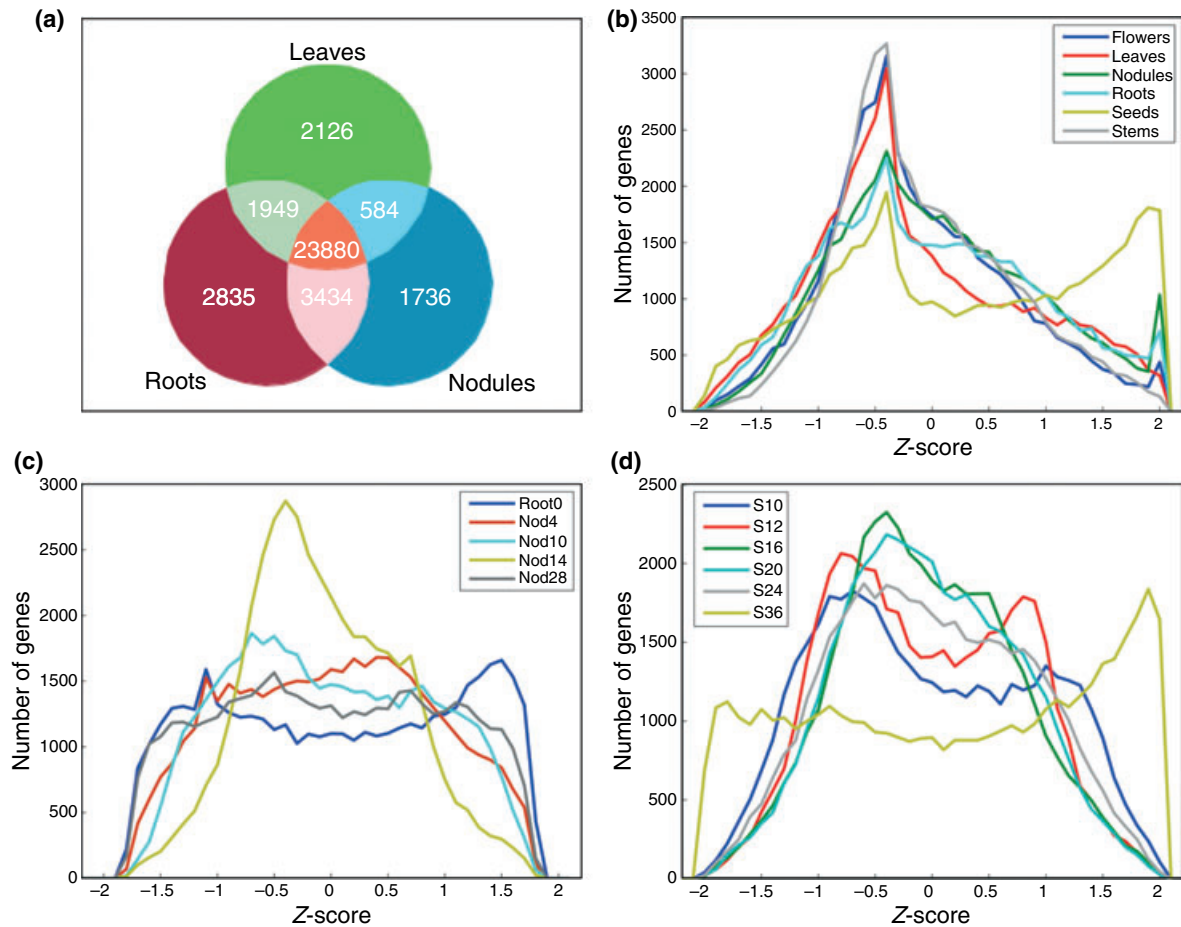
Transcripts were detected in at least one organ for 86.1% of all genes represented by the 50 900 probe sets on the Medicago Genome Chip. There was a high degree of overlap in the sets of genes expressed in different organs: Transcripts of 42% of all expressed genes (36% of all probe sets; Figure S1) were detected in all organs, and this percentage increased in pair-wise comparisons between organs. For example, 79% of genes expressed in roots or nodules of 28-day-old plants were expressed in both organs (Figure 1a). Despite the qualitative similarities in gene expression amongst organs, the dynamics of gene expression differed markedly between organs and within organs over developmental time (Figure 1b–d). The majority of genes were subject to transcriptional and/or post-transcriptional regulation that altered transcript levels during plant development. In fact, at least 73% of all genes exhibited a >100% change in transcript level from the organ with lowest expression to the organ with highest expression. The mean coefficient of variance (CV = standard deviation/mean) of transcript levels for all expressed genes across all organs was 60.6%, ranging from 2.3% to 428.6%, while the mean CV for the three biological replicates of each organ was only 13.3%. In other words, the biological variation in gene transcript levels within an organ, including technical errors associated with measurement, was far less than the biological variation between organs.

Similarity between the transcriptomes of different organs was estimated using Pearson correlation, taking into account all genes expressed in at least one organ. The resulting heatmap of correlations revealed three main groups of organs with similar transcriptomes (Figure 2). The first group consisted of the underground organs, roots and nodules, the second group included seeds at different developmental stages, and the third group contained aerial organs, including leaves, stems, petioles, shoot apices, pods, and flowers. The aerial organs were more closely related to seeds than to underground organs.

Consistent with the large degree of overlap between expressed genes in the various organs, relatively few genes were found to be expressed in an organ-specific manner (Figure 3). The number of organ-specific genes identified in samples from 4-week-old plants ranged from 10 for stems to 322 for nodules. These numbers increased when data were obtained from multiple stages during development, as







**Figure 1.** Gene expression dynamics during *Medicago* development.

(a) Comparison of gene expression (detected transcripts) in roots, nodules, and leaves of 28-day-old plants.

(b–d) Relative gene expression levels (Z scores) in different organs at one developmental stage (b), and at multiple stages for nodules (c) and seeds (d). Transcript levels were  $\log_2$ -transformed before calculation of  $Z = (X - X_{av})/SD$ ; where  $X$  is the mean transcript level for a given gene in the specified organ,  $X_{av}$  is the average transcript level for that gene in all organs, and  $SD$  is the standard deviation of transcript level for that gene across all organs. The number of genes was determined for each  $\Delta Z = 0.1$ . Nod4–Nod28 represent nodules harvested 4–28 days after inoculation with *Sinorhizobium meliloti*, while Root0 represents root tissue immediately prior to inoculation. S10–S36 represent seed harvested from 10 to 36 days after pollination.

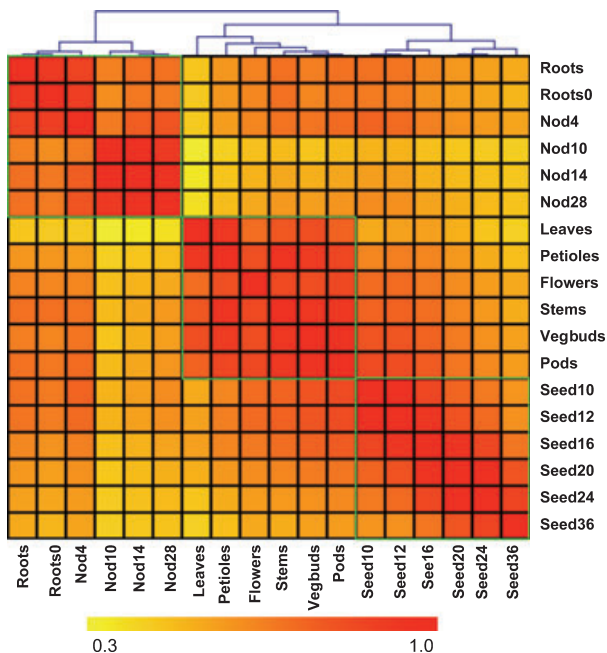
exemplified by nodules and seeds in which 473 and 584 genes, respectively, were identified as organ-specific from the corresponding developmental series (Table S1). In some cases, organ-specific genes were expressed only transiently during development, suggesting to us that they play roles in development *per se*, rather than in the maintenance of specialized biochemical or physiological functions of each organ, at least under ideal growth conditions.

A more complex picture emerged for genes that were expressed in specific organs at a level at least twice that of any other organ (Figure S2). The number of such genes ranged from 40 in petioles to 908 in roots of 4-week-old plants. Even larger numbers of organ-induced genes were uncovered in the developmental time-course for nodules (1354 genes) and seeds (3228 genes; Table S2). Interestingly, transcript levels of many of the genes induced during nodule or seed development were maintained at relatively high levels in the mature organ, suggesting that they may

play roles in differentiation or the maintenance of specialized organ functions.

To identify all genes that are subject to transcriptional or post-transcriptional regulation during the development of *Medicago*, we chose roots as an arbitrary reference organ and tested the null hypothesis that expression in other organs was not significantly different from that in roots. Using a Bonferroni-corrected (Abdi, 2007)  $P$ -value threshold of  $1.14 \times 10^{-6}$ , the percentage of genes expressed at a different level from roots ranged from 46.7% in nodules to 55.9% in leaves. In total, 81.5% of genes were differentially expressed between roots and one or more of the other seven organs (Table S3). The false discovery rate of these genes was estimated by determining  $Q$ -values for each, using EDGE software (Leek *et al.*, 2006; Storey and Tibshirani, 2003). Clearly, organ development, differentiation, and maintenance in plants are underpinned by massive quantitative changes in gene expression.



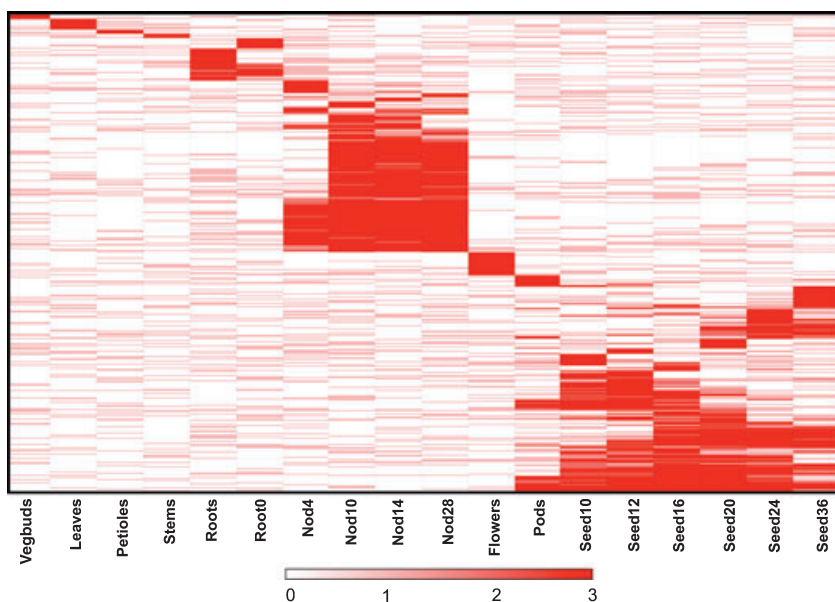


**Figure 2.** Comparison of transcriptomes of various organs. Pair-wise Pearson correlation coefficients were used to generate the heat map. The color scale indicates the degree of correlation. Samples were clustered with Euclidean distance using the MultiExperiment Viewer (MeV, <http://www.tm4.org/mev.html>).

Genes with a constant expression level throughout development and in the face of environmental challenges, which often fulfill housekeeping roles in cells, are useful reference points for comparative gene expression analysis (Czechowski *et al.*, 2005). We identified 102 genes with <16% coefficient of variance for transcript levels amongst all the

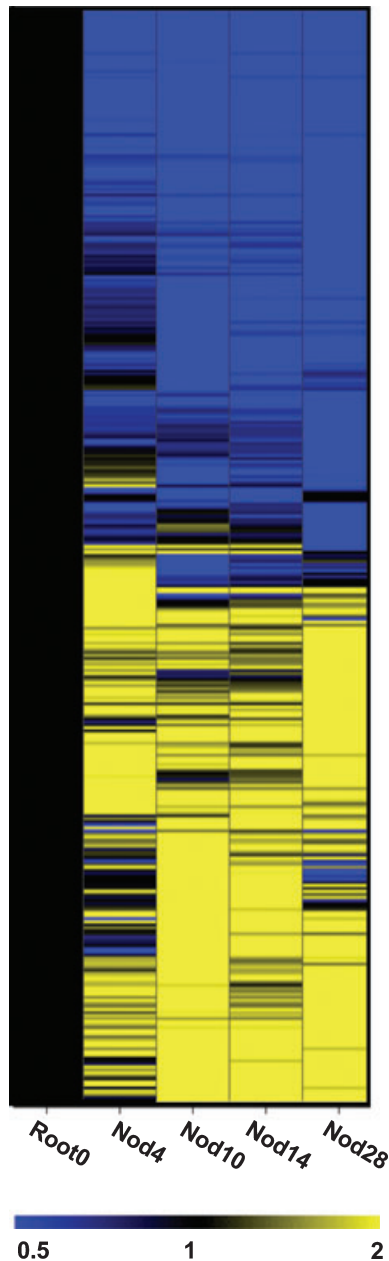
organs analyzed (Table S4). Transcript levels of these genes ranged widely, from as high as 14 500 to as low as 100 units, which was used as the minimum threshold level. Amongst the stably expressed genes were several that are used traditionally as reference genes in plants, including glyceraldehyde-3-P-dehydrogenase and ubiquitin. The most stably expressed gene, corresponding to TC97716 with unknown function, had a CV for transcript level amongst all organs of 9.4%. This set of reference genes will be of great value to legume research for normalizing gene expression data derived from qRT-PCR or probe-hybridization approaches.

Symbiotic nitrogen fixation in plants is a process confined largely to the legume (Fabaceae) family. Therefore, well-established, non-legume model species such as *Arabidopsis thaliana* and *Oryza sativa* (rice) cannot be used to learn more about SNF. The data presented here represent the most comprehensive data set to date for gene expression during nodule development in a legume. More than 26 000 genes are expressed during nodule development and 30.2% of these are differentially expressed (transcript levels increase or decrease more than twofold compared with roots with Bonferroni-corrected  $P < 1.14 \times 10^{-6}$ ) at some stage during this development (Figure 4 and Table S5). Visualization of the nodule development data, using MAPMAN to overlay changes in gene expression onto metabolic maps (Goffard and Weiller, 2006; Thimm *et al.*, 2004), confirmed and extended previous, smaller-scale transcriptomics studies (El-Yahyaoui *et al.*, 2004) that showed induction during nodule development of genes involved in glycolysis, carbon fixation, and nitrogen metabolism (Figure S3a–c). Many genes involved in secondary metabolism, such as the terpenoid and flavonoid pathways, were repressed during nodule development (Figure S3d). These results are



**Figure 3.** Heat map of organ-specific genes. The color scale indicates the number of times transcripts for a given gene were detected in the three biological replicates of each organ. Only those genes are shown for which transcripts were detected in all three biological replicates of one organ and no more than once in another organ.





**Figure 4.** Hierarchical clustering of genes that were differentially expressed during nodule development. Clustering of genes was based on Pearson correlation. The heat map portrays transcript levels in nodules 4, 10, 14, and 28 days after inoculation (Nod4–Nod28) relative to that in uninoculated roots (Root0).

interesting in light of the fact that many secondary compounds play roles in plant defense (Dixon, 2001), a process that is presumably suppressed in nodules in order to maintain a quasi-stable symbiosis with the nitrogen-fixing rhizobia. We provide the nodule development data described here in a form that can be imported into MAPMAN for exploration by the reader of other pathways and processes (Table S6).

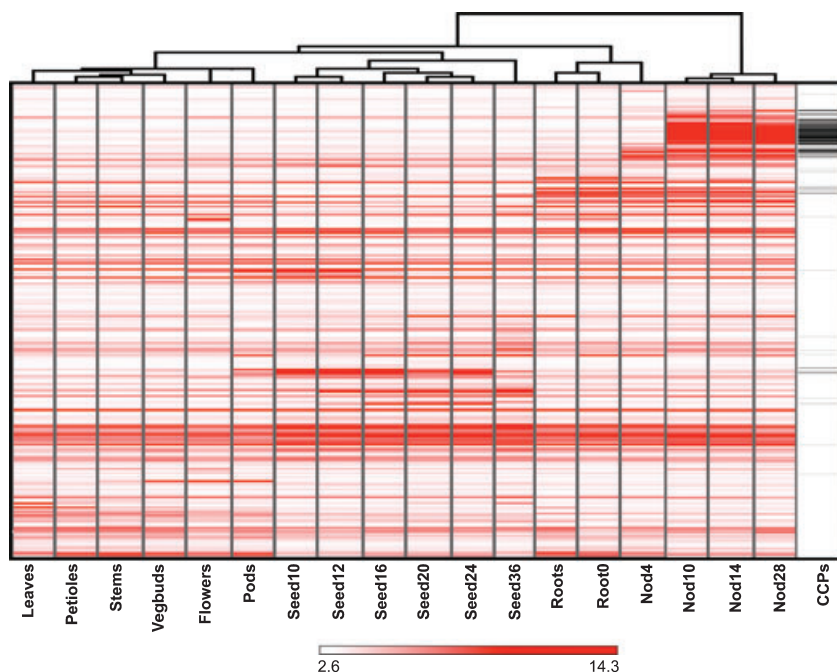
The massive changes in transcript abundance that occur during the development of nodules and other organs indicate tremendous regulatory activity at the transcriptional and/or post-transcriptional levels. Transcription factors (TFs) are DNA-binding proteins that interact with specific *cis*-elements of genes to regulate transcription, either positively or negatively. Plants such as *Arabidopsis* may possess as many as 2000 TF genes, representing more than 6% of all their genes (Riechmann and Ratcliffe, 2000; Riechmann *et al.*, 2000). To identify TF genes that control development and differentiation in *Medicago*, we first created a list of putative TF genes by screening predicted protein sequences for the presence of known or suspected DNA-binding domains, using InterPro (<http://www.ebi.ac.uk/interpro/>) and Pfam (<http://pfam.sanger.ac.uk/>) domain identification, additional hidden Markov model (HMM) predictions (Guo *et al.*, 2005; Sonnhammer *et al.*, 1997), and a BLASTX search of the NCBI NR database to support annotations. In this way, we previously identified 1298 putative TF genes represented by probe sets on the Affymetrix *Medicago* GeneChip (Udvardi *et al.*, 2007). Most of these (1169) fall into the 45 known families of plant TF genes or other transcriptional regulators, while 129 may define novel TF families in plants (Table S7). Five hundred and thirty-two of the putative TF genes are differentially expressed (more than twofold change;  $P < 1.14 \times 10^{-6}$ ) during nodule development and may therefore play important roles in SNF (Table S5). The vast majority (>1100) of TFs are differentially expressed in other organs (Table S3). These data are a rich source of information and a sound platform for future experimental work aimed at unraveling genetic regulatory networks that govern organ development in *Medicago*.

The ability to form a nitrogen-fixing symbiosis appears to have evolved relatively recently in land plants, approximately 65 million years ago (Doyle, 1998), and as a result SNF is restricted to legumes and a few non-legume species. Interestingly, some of the genes required to establish SNF in legumes appear to have been recruited from a more ancient set of genes that are required for mycorrhizal symbiosis (Kistner and Parniske, 2002). Mycorrhizal symbioses are believed to have evolved when plants first colonized land 450 million years ago (Redecker *et al.*, 2000; Remy *et al.*, 1994), and it has been suggested that these fungal symbionts served as an extension of the primitive plant root system. Indeed, mycorrhizal fungi extend the reach of plant roots and aid in plant nutrition, especially phosphorus uptake (Harrison, 1999). The ancient origin and importance of mycorrhizal symbioses to land plants is reflected by the fact that approximately 90% of all land plant species are able to form such symbioses. A number of genes, mostly encoding signaling proteins, have been discovered in legumes that are required for both mycorrhizal symbiosis and nodulation/SNF (Kistner and Parniske, 2002; Parniske, 2004). Additional genes are



essential for SNF but not for mycorrhizal symbiosis (Kalo *et al.*, 2005; Radutoiu *et al.*, 2003; Schauser *et al.*, 1999; Smit *et al.*, 2005). Each of these additional genes has so far been found to have one or more homologs in non-legume, non-nitrogen-fixing plant species, such as Arabidopsis. This raises an important question: Were all genes required for nodule development and SNF simply recruited from a pre-existing stock of plant genes, or did novel genes evolve as a result of natural selection for SNF? Legume-specific genes (LSGs) that appear to be absent from the genomes of non-legumes have been identified in a number of species. Several classes of LSGs have been identified in Medicago that encode short proteins, including over 300 cysteine cluster proteins (CCPs; Alunni *et al.*, 2007; Fedorova *et al.*, 2002; Graham *et al.*, 2004; Mergaert *et al.*, 2003), 63 proline-rich proteins (PRPs; Graham *et al.*, 2004; Sherrier *et al.*, 2005) and 21 glycine-rich proteins (GRPs; Alunni *et al.*, 2007; Kevei *et al.*, 2002; Silverstein *et al.*, 2006). Five thousand eight hundred and forty-two probe sets representing LSGs were identified on the Medicago GeneChip. This included sequences for 355 CCPs, 5 PRPs, and 7 GRPs (Table S8). Analysis of LSG expression revealed that a subset of 322 CCPs (called NCRs for nodule-specific cysteine rich) and all seven GRPs were expressed in a nodule-specific manner consistent with roles for these genes in nodule development and/or function (Figure 5 and Table S8). Some PRPs were expressed in nodules but were also detected in other tissues such as seeds or flowers (Figure 5). These results confirm and extend earlier work on LSGs (Alunni *et al.*, 2007; Fedorova *et al.*, 2002; Graham *et al.*, 2004; Mergaert *et al.*, 2003).

Large data sets of the type presented here for Medicago and elsewhere for different plant species, such as Arabidopsis (Schmid *et al.*, 2005), enable us to address other important questions about gene evolution in plants. One of these questions relates to the conservation of gene function in different plant lineages. To gain insight into the possible role(s) of a gene in a crop species, for instance, plant biologists often turn to a related model species, such as Arabidopsis, rice, or Medicago, and ask: what is the function of the ortholog(s) of the crop gene(s)/protein(s) in the model species? The implicit assumption is that orthologs in different species perform similar, if not identical, physiological functions despite millions of years of evolution. For this to be true, orthologous genes must have similar expression profiles in the two organisms. To test whether this is generally the case, we identified homologs of Medicago genes in Arabidopsis and made pair-wise comparisons of gene expression between the two species, using matching data from roots, stems, leaves, flowers, seeds, petioles, and vegetative buds. Pearson correlation analysis was used to rank Arabidopsis homologs based on the similarity of gene expression profiles in the various organs of the two species. The top-ranking homolog determined from correlation analysis of gene expression matched the putative ortholog, based on phylogenetic analysis in only 62% of cases (Figure 6 and Table S9). Thus, transcriptional regulation of putative orthologs has diverged substantially between Medicago and Arabidopsis, indicating that many orthologs may not perform the same range of functions in these two species. It should be noted that the plant growth conditions and developmental states of Medicago and Arabidopsis



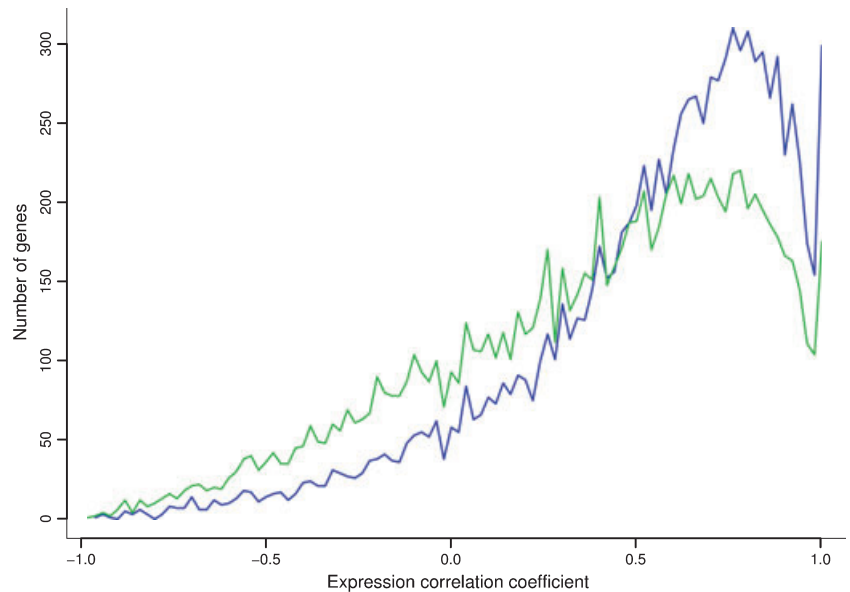
**Figure 5.** Expression of legume-specific genes in Medicago organs.

Genes were clustered based on Pearson correlation. Cysteine cluster proteins (CCPs) are indicated in the far-right column. The color scale shows  $\log_2$ -transformed transcript levels for each gene.





**Figure 6.** Correlation between expression profiles of homologous and orthologous genes in *Arabidopsis thaliana* and *Medicago truncatula*. Normalized transcript levels of roots, stems, leaves, flowers, seeds, petioles, and vegetative buds were compared between the two species. Pearson (linear) correlation coefficients were determined for all pairs of sequences regarded as either homologous or orthologous between *Medicago* and *Arabidopsis*. Histograms of the number of sequences (Y-axis) over the correlation coefficient (X-axis) of *Medicago* sequences and the corresponding *Arabidopsis* sequences are shown. The blue line represents the best-correlating *Arabidopsis* sequence identifiable within the set of sequence homologs for each *Medicago* sequence. The green line represents the best-correlating *Arabidopsis* sequence identifiable within the set of putative sequence orthologs for each *Medicago* sequence.



organs compared here were not identical, which would tend to decrease the correlation between gene expression patterns in the two species. Nonetheless, given that the expression patterns of paralogs often exhibited higher correlation than predicted orthologs (Figure 6), phylogenetic analysis alone may yield inaccurate predictions for the physiological functions of many genes, at least for comparisons between families as divergent as legumes and crucifers. This underscores the importance of the gene expression data collected here as a tool for *Medicago* and legume functional genomics.

In summary, we have produced a comprehensive gene expression atlas for the model legume *M. truncatula*, which encompasses all organs of this species, including detailed time-courses through nodule and seed development. In addition to being a rich source of information for legume biologists, this data set enables large-scale comparisons between the transcriptomes of different plant species. Analysis of the data presented here shows that differences between plant organs result mainly from quantitative rather than qualitative changes in global gene expression. Relatively few genes are organ-specific in this species. Amongst these are subsets of legume-specific genes that appear to be expressed exclusively or preferentially in nodules. This implies that evolution of symbiotic nitrogen fixation was accompanied, and possibly facilitated, by the evolution of novel genes in legumes. These genes are clear targets for future studies aimed at identifying the core set of genes required for SNF in legumes. Finally, comparative functional genomics studies of the type presented here for *Medicago* and *Arabidopsis* will add a new dimension to studies of gene evolution in plants and other organisms.

We plan to expand the *Medicago* Gene Expression Atlas to encompass transcript data from wild-type and mutant plants exposed to various biotic and abiotic challenges, and to increase the spatial resolution of expression data by analyzing specific tissues and cell types. We invite the scientific community to collaborate with us in this venture by submitting raw *Medicago* Affymetrix data from complementary experiments together with metadata describing the experimental material, either directly to us (contact [mudvardi@noble.org](mailto:mudvardi@noble.org)) or to ArrayExpress (<http://www.ebi.ac.uk/miamexpress/>).

## Experimental procedures

### *Plant material, RNA isolation, probe preparation and array hybridization*

*Medicago truncatula* cv. Jemalong, line A17 seeds were scarified with concentrated sulfuric acid, rinsed, sterilized with 2% sodium hypochlorite, and vernalized at 4°C for 3 days on moist, sterile filter paper. Germinated seedlings were transplanted to pots containing Surface MVP calcined (illite) clay (Profile Products, <http://www.profileproducts.com/>) and placed in a growth chamber set to the following conditions: 16-h/8-h light/dark regime, 200  $\mu\text{E m}^{-2} \text{sec}^{-1}$  light irradiance, 24°C and 40% relative humidity. Plants were fertilized daily with half-strength B&D solution (Broughton and Dilworth, 1971) containing 2 mM  $\text{KNO}_3$  and 2 mM  $\text{NH}_4\text{NO}_3$ . A subset of plants were inoculated with *Sinorhizobium meliloti* strain 1021 at 1 and 7 days after sowing and fertilized with half-strength B&D solution containing 0.5 mM  $\text{KNO}_3$ . Vegetative organs (roots, stems, petioles, leaves, and shoot buds from uninoculated plants and nodules from inoculated plants) were harvested from multiple plants 28 days after planting and pooled for individual biological replicates in a completely randomized design. All experiments were performed with three biological replicates planted on separate days. For flowers and pods, plants were vernalized for 2 weeks to decrease



time to flowering. Flowers were harvested on the first day that they opened fully, while pods were collected at various stages of development (length ranged from 2.5 to 9.0 mm) within 21 days after the appearance of the floral bud. Harvesting of all organs occurred at the same time each morning, approximately 3 h after 'dawn', to avoid as far as possible diurnal changes in gene expression. All harvested material was frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  prior to RNA isolation. Total RNA was extracted using TRIzol reagent (Invitrogen, <http://www.invitrogen.com/>; Chomczynski and Mackey, 1995), treated with DNaseI (Ambion, <http://www.ambion.com/>), and column purified with a RNeasy MinElute CleanUp Kit (Qiagen, <http://www.qiagen.com/>).

Material for the nodule developmental series was harvested from plants grown aeroponically at  $22^{\circ}\text{C}$ , 75% hygrometry, a light irradiance of  $200\ \mu\text{E m}^{-2}\ \text{sec}^{-1}$ , and a 16-h/8-h light/dark regime. Plants were grown initially for 11 days using a nitrogen-rich medium (Journet *et al.*, 2001) then deprived of nitrogen for 4 days before being inoculated with *S. meliloti* strain 2011. Roots were harvested 0, 4, 10, and 14 days post-inoculation and nodules were dissected from roots prior to freezing in liquid nitrogen, storage at  $-80^{\circ}\text{C}$ , and RNA isolation using a Nucleospin RNA kit (Macherey-Nagel, <http://www.macherey-nagel.com/>).

Material for the seed developmental series was harvested from plants grown in pots containing attapulgate (50%) and clay beads (50%) at  $22^{\circ}\text{C}/19^{\circ}\text{C}$  day/night, 16-h photoperiod at  $220\ \mu\text{E m}^{-2}\ \text{sec}^{-1}$  light irradiance, and 60–70% relative humidity. Plants were fertilized with nutrient solution three times a week and watered on intervening days. Seeds were excised from pods 10, 12, 16, 20, 24, and 36 days after pollination, frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$  prior to RNA isolation (Chang *et al.*, 1993).

Ribonucleic acid was quantified using a Nanodrop Spectrophotometer ND-100 (NanoDrop Technologies, <http://www.nanodrop.com/>) and evaluated for purity with a Bioanalyzer 2100 (Agilent, <http://www.home.agilent.com/>). The Affymetrix GeneChip<sup>®</sup> Medicago Genome Array (Affymetrix, <http://www.affymetrix.com/>) was used for expression analysis. The RNA from three independent biological replicates was analyzed for each organ/developmental stage. Probe labeling using  $10\ \mu\text{g}$  RNA, array hybridization and scanning were performed according to the manufacturer's instructions (Affymetrix) for eukaryotic RNA, using a one-cycle protocol for cDNA synthesis.

#### Data extraction and normalization

For each Affymetrix array hybridized, the resulting .cel file was exported from GeneChip Operating Software Version 1.4 (Affymetrix) and imported into Robust Multiarray Average (RMA; Irizarry *et al.*, 2003) for global normalization. Presence/absence call for each probe set was obtained using dCHIP (Li and Wong, 2001). Gene selections based on an associative *t*-test (Dozmorov and Centola, 2003) were made using Matlab (MathWorks, <http://www.mathworks.com/>). Using this method, the background noise presented between replicates and the technical noise generated during hybridization were measured by the residual presented among a group of genes whose residuals are homoscedastic within the control group. Only genes whose residuals between compared sample pairs are significantly higher than the measured noise level will be considered to be differentially expressed. Since the residual was obtained from thousands of genes on the chip, the *P*-values obtained by this method are corrected towards a large sampling size, thus enabling the use of Bonferroni corrections without being overly stringent. The advantage of this methodology is that it takes into consideration technical noise and internal variations between replicates within a sample group and provides a baseline for

selecting biologically significant genes. A selection threshold of two for transcript ratios (where applicable) and a Bonferroni-corrected *P*-value threshold of  $1.14 \times 10^{-6}$  were used. Bonferroni-corrected  $P = 0.05/N$ , where *N* is the number of genes in the comparison, which was 43 836 in the experiments reported here. To monitor the false discovery rate of differentially expressed genes, the *Q*-value of each gene was obtained by EDGE software (Leek *et al.*, 2006; Storey and Tibshirani, 2003).

#### Z scores

The Z score was calculated as follows:  $Z = (X - X_{\text{av}})/SD$ ; where *X* is the  $\log_2$ -transformed mean transcript level for a given gene in a specific organ,  $X_{\text{av}}$  is the  $\log_2$ -transformed mean transcript level for that gene in all organs, and SD is the standard deviation of transcript level for that gene across all organs. Mean transcript levels were determined from three biological replicates of each organ.

#### Hierarchical clustering analysis (HCA)

Hierarchical clustering analysis was conducted with Spotfire DecisionSite 8.1 (Spotfire Inc., <http://spotfire.tibco.com/>). For clustering analysis of data from different organs, data were transformed into  $\log_2$  and clustered using Pearson correlation analysis (Zar, 1999). For the nodule developmental series, transcript levels were expressed relative to the level in roots just prior to inoculation with rhizobia (Root 0) before constructing clusters using the Pearson correlation coefficient.

#### Legume-specific genes

Legume-specific genes (LSGs) represented on the Affymetrix Medicago GeneChip were identified by a series of in-house BLAST searches that were used to eliminate probe sets representing sequences with homology to any non-legume plant sequence in GenBank. Briefly, starting with all 50 900 sequences upon which the Medicago GeneChip was based, consecutive BLAST searches were used to filter out homologs from *O. sativa*, *A. thaliana*, *Populus trichocarpa*, and *Chlamydomonas reinhardtii*. Finally, using the reduced list, a final search against all remaining non-legume sequences in GenBank NR and EST databases was made using the BLAST Network Client 'BLASTcl3'. (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/>). For all these searches both TBLASTX and BLASTN (*E*-value cutoff of  $\leq 1 \times 10^{-4}$ ) were used. Subsets of LSGs (CCPs and GRPs) represented on the Affymetrix Medicago GeneChip were identified based on homology to known family members using a TBLASTX search (match criteria: *E*-value  $\leq 1 \times 10^{-5}$ ). Probe sets corresponding to known PRPs were identified based on perfect or near perfect matches from a BLASTN search employing a complexity filter.

#### Correlation analysis of expression profiles for homologous genes in Medicago and Arabidopsis

All *M. truncatula* consensus sequences (<http://www.affymetrix.com/support/technical/byproduct.affx?product=medicago>), the sequences from which the probe sets for the corresponding Affymetrix chip were derived, were translated into protein in all six frames. The longest open reading frame (ORF) for each sequence was compared, using BLASTP, against six-frame translations of the consensus sequences of other Affymetrix chips as well as the NCBI non-redundant sequence database 'nr'. Consensus sequences for the *M. truncatula* and Arabidopsis chips were accepted as sequence



homologs if BLASTP hits connected the sequences with *E*-values better than  $1 \times 10^{-30}$  and bit-scores greater than 150. Putative phylogenetic orthologs between *M. truncatula* and *A. thaliana* were identified using AffyTrees (Frickey *et al.*, 2008). AffyTrees is based on PhyloGenie (Frickey and Lupas, 2004) and provides a repository of neighbor-joining (Saitou and Nei, 1987) trees for Affymetric consensus sequences in plants.

To compare the gene expression of homologous sequences, *A. thaliana* microarray data for stems (ATGE\_27), petioles (ATGE\_19), leaves (ATGE\_14), vegetative buds (ATGE\_8), flowers (ATGE\_39), roots (ATGE\_9), and seeds (ATGE\_79; Schmid *et al.*, 2005) were compared against the corresponding organs provided by this atlas (stem, petiole, leaf, vegetative bud, flower, root, and seed20d). All expression data were normalized using GCRMA (Wu *et al.*, 2004). The Pearson (linear) correlation coefficient of the expression values was calculated for all pairs of sequence homologs and used to quantify the similarity of expression of homologous sequences for the two species. As the number of *Medicago* consensus sequences for which BLAST and AffyTrees could determine homologs or orthologs differed, we restricted the analysis to the 10 243 sequences for which both methods were able to produce results. Sequence pairs between *M. truncatula* and *A. thaliana* with the highest correlation coefficient within the set of sequence homologs, as determined by BLASTP, and the set of putative orthologs, as determined by AffyTrees, were compared.

Our project web site is <http://bioinfo.noble.org/gene-atlas/>. All gene expression data have been deposited in the ArrayExpress Database (<http://www.ebi.ac.uk/miameexpress/>) under accession number E-MEXP-1097.

### Acknowledgements

We thank the USDA CSREES-NRI, the Samuel Roberts Noble Foundation, the Max Planck Society, the European Union FP6 Program, and the Australian Research Council Centre for Integrative Legume Research for support of this work.

### Supplementary Material

The following supplementary material is available for this article online:

**Figure S1.** Fraction of genes expressed in different organs.

**Figure S2.** Heat map of organ-induced genes. Transcripts levels for each of these genes were at least twice as high in one organ as in any other organ.

**Figure S3.** Transcriptional dynamics during nodule development of genes encoding enzymes involved in: (a) glycolysis, (b) carbon fixation, (c) nitrogen metabolism; and (d) flavonoid biosynthesis.

**Table S1.** List of organ-specific genes and their presence calls in all organs.

**Table S2.** List of organ-induced genes and their transcript levels in all organs.

**Table S3.** Differentially expressed genes in *Medicago*.

**Table S4.** Stably expressed 'reference' genes for transcript normalization in *Medicago*.

**Table S5.** Genes differentially expressed during nodule development.

**Table S6.** Nodule developmental data for visualization with MAPMAN software.

**Table S7.** Differential expression of transcription factor genes in various organs of *Medicago*.

**Table S8.** Legume-specific genes and their expression level in different organs.

**Table S9.** Information about the various BLAST hits (better than  $1 \times 10^{-30}$  and bit-scores of 150) as well the sequences present in the AffyTrees phylogenies and the corresponding Pearson (linear) correlation of the expression values.

This material is available as part of the online article from <http://www.blackwell-synergy.com>.

Please note: Blackwell publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

### References

- Abdi, H. (2007) Bonferroni and Sidak corrections for multiple comparisons. In *Encyclopedia of Measurement and Statistics* (Salkind, N.J., ed.). Thousand Oaks: Sage, pp. 103–107.
- Alunni, B., Kevei, Z., Redondo-Nieto, M., Kondorosi, A., Mergaert, P. and Kondorosi, E. (2007) genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. *Mol. Plant Microbe Interact.* **20**, 1138–1148.
- Becker, A., Berges, H., Krol, E. *et al.* (2004) Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Mol. Plant Microbe Interact.* **17**, 292–303.
- Brewin, N.J. (1991) Development of the legume root nodule. *Annu. Rev. Cell Biol.* **7**, 191–226.
- Broughton, W.J. and Dilworth, M.J. (1971) Control of leghaemoglobin synthesis in Snake Beans. *Biochem. J.* **125**, 1075–1080.
- Chang, S., Puryear, J. and Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**, 113–116.
- Chomczunski, P. and Mackey, K. (1995) Modification of the TRI TM Reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. *Biotechniques*, **19**, 942–945.
- Colebatch, G., Trevaskis, B. and Udvardi, M. (2002) Symbiotic nitrogen fixation research in the postgenomics era. *New Phytol.* **153**, 37–42.
- Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J. and Udvardi, M.K. (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. *Plant J.* **39**, 487–512.
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. and Scheible, W.R. (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17.
- Dixon, R.A. (2001) Natural products and plant disease resistance. *Nature*, **411**, 843–847.
- Doyle, J.J. (1998) Phylogenetic perspectives on nodulation: an evolving view of plants and symbiotic bacteria. *Trends Plant Sci.* **3**, 473–478.
- Dozmorov, I. and Centola, M. (2003) An associative analysis of gene expression array data. *Bioinformatics*, **19**, 204–211.
- El-Yahyaoui, F., Küster, H., Amor, B.B. *et al.* (2004) Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. *Plant Physiol.* **136**, 3159–3176.
- Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J., Town, C.D., VandenBosch, K.A., Gantt, J.S. and Vance, C.P. (2002) Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. *Plant Physiol.* **130**, 519–537.
- Frickey, T. and Lupas, A.N. (2004) PhyloGenie: automated phylogeny generation and analysis. *Nucleic Acids Res.* **32**, 5231–5238.



- Frickey, T., Benedito, V.A., Udvardi, M. and Weiller, G. (2008) AffyTrees: facilitating comparative analysis of Affymetrix plant microarray chips. *Plant Physiol.* **146**, 377–386.
- Goffard, N. and Weiller, G. (2006) Extending MapMan: application to legume genome arrays. *Bioinformatics*, **22**, 2958–2959.
- Graham, P.H. and Vance, C.P. (2003) Legumes: importance and constraints to greater use. *Plant Physiol.* **131**, 872–877.
- Graham, M.A., Silverstein, K.A.T., Cannon, S.B. and VandenBosch, K.A. (2004) Computational identification and characterization of novel genes from legumes. *Plant Physiol.* **135**, 1179–1197.
- Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L. and Luo, J. (2005) DATF: a database of Arabidopsis transcription factors. *Bioinformatics*, **21**, 2568–2569.
- Harrison, M.J. (1999) Molecular and cellular aspects of the arbuscular mycorrhizal symbiosis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 361–389.
- Irizarry, R.A., Hobbs, B. and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.
- Journet, E.P., El-Gachtouli, N., Vernoud, V., de Billy, F., Pichon, M., Dedieu, A., Arnould, C., Morandi, D., Barker, D.G. and Gianinazzi-Pearson, V. (2001) Medicago truncatula ENOD11: a novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells. *Mol. Plant Microbe Interact.* **14**, 737–748.
- Kalo, P., Gleason, C., Edwards, A. et al. (2005) Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science*, **308**, 1786–1789.
- Kevei, Z., Vinardell, J.M., Kiss, G.B., Kondorosi, A. and Kondorosi, E. (2002) Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in Medicago spp. *Mol. Plant Microbe Interact.* **15**, 922–931.
- Kistner, C. and Parniske, M. (2002) Evolution of signal transduction in intracellular symbiosis. *Trends Plant Sci.* **7**, 511–518.
- Leek, J.T., Monsen, E., Dabney, A.R. and Storey, J.D. (2006) EDGE: extraction and analysis of differential gene expression. *Bioinformatics*, **22**, 507–508.
- Li, C. and Wong, W. (2001) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA*, **98**, 31–36.
- Long, S.R. (2001) Genes and signals in the Rhizobium-legume symbiosis. *Plant Physiol.* **125**, 69–72.
- Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A. and Kondorosi, E. (2003) A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* **132**, 161–173.
- Mitra, R.M., Shaw, S.L. and Long, S.R. (2004) Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. *Proc. Natl Acad. Sci. USA*, **101**, 10217–10222.
- Oldroyd, G.E.D. and Downie, J.A. (2004) Calcium, kinases, and nodulation signalling in legumes. *Nat. Rev. Mol. Cell Biol.* **5**, 566–576.
- Parniske, M. (2004) Molecular genetics of the arbuscular mycorrhizal symbiosis. *Curr. Opin. Plant Biol.* **7**, 414–421.
- Radutoiu, S., Madsen, L.H., Madsen, E.B. et al. (2003) Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, **425**, 585–592.
- Redecker, D., Kodner, R. and Graham, L.E. (2000) Glomalean fungi from the Ordovician. *Science*, **289**, 1920–1921.
- Remy, W., Taylor, T.N., Hass, H. and Kerp, H. (1994) Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Natl Acad. Sci. USA*, **91**, 11841–11843.
- Riechmann, J.L. and Ratcliffe, O.J. (2000) A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.* **3**, 423–434.
- Riechmann, J.L., Heard, J., Martin, G. et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, **290**, 2105–2110.
- Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Schauser, L., Roussis, A., Stiller, J. and Stougaard, J. (1999) A plant regulator controlling development of symbiotic root nodules. *Nature*, **402**, 191–195.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D. and Lohmann, J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**, 501–506.
- Sherrier, D.J., Taylor, G.S., Silverstein, K.A.T., Gonzales, M.B. and VandenBosch, K.A. (2005) Accumulation of extracellular proteins bearing unique proline-rich motifs in intercellular spaces of the legume nodule parenchyma. *Protoplasma*, **225**, 43–55.
- Silverstein, K.A.T., Graham, M.A. and VandenBosch, K.A. (2006) Novel paralogous gene families with potential function in legume nodules and seeds. *Curr. Opin. Plant Biol.* **9**, 142–146.
- Smil, V. (1999) Nitrogen in crop production: an account of global flows. *Global Biogeochem. Cycles*, **13**, 647–662.
- Smit, P., Raedts, J., Portyanko, V., Debelle, F., Gough, C., Bisseling, T. and Geurts, R. (2005) NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science*, **308**, 1789–1791.
- Sonnhammer, E.L.L., Eddy, S.R. and Durbin, R. (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins*, **28**, 405–420.
- Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genome wide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y. and Stitt, M. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939.
- Town, C.D. (2006) Annotating the genome of *Medicago truncatula*. *Curr. Opin. Plant Biol.* **9**, 122–127.
- Uchiumi, T., Ohwada, T., Itakura, M. et al. (2004) Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *J. Bacteriol.* **186**, 2439–2448.
- Udvardi, M.K. and Day, D.A. (1997) Metabolite transport across symbiotic membranes of legume nodules. *Annu. Rev. Plant Physiol. Plant Molec. Biol.* **48**, 493–523.
- Udvardi, M.K., Kakar, K., Wandrey, M. et al. (2007) Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol.* **144**, 538–549.
- Wu, Z., Irizarry, R.A., Gentleman, R., Murillo, F.M. and Spencer, F. (2004) *A Model Based Background Adjustment for Oligonucleotide Expression Arrays*. Technical Report. Baltimore, MD: John Hopkins University, Department of Biostatistics Working Papers.
- Zabala, G., Zou, J., Tuteja, J., Gonzalez, D.O., Clough, S.J. and O., V.L. (2006) Transcriptome changes in the phenylpropanoid pathway of *Glycine max* in response to *Pseudomonas syringae* infection. *BMC Plant Biol.* **6**, 26.
- Zar, J.H. (1999) *Biostatistical Analysis*. Upper Saddle River, NJ: Prentice Hall.



GO:0008150 biological_process (biological_process)	4867
GO:0009653 anatomical structure morphogenesis (anatomical structure morphogenesis)	5
GO:0016049 cell growth (cell growth)	0
GO:0007610 behavior (behavior)	0
GO:0009987 cellular process (cellular process)	3320
GO:0006519 amino acid and derivative metabolic process	173
GO:0007154 cell communication (cell communication)	181
GO:0007267 cell-cell signaling (cell-cell signaling)	0
GO:0009875 pollen-pistil interaction (pollen-pistil interaction)	0
GO:0009991 response to extracellular stimulus	0
GO:0007165 signal transduction (signal transduction)	173
GO:0007049 cell cycle (cell cycle)	29
GO:0030154 cell differentiation (cell differentiation)	33
GO:0008219 cell death (cell death)	31
GO:0019725 cell homeostasis (cell homeostasis)	54
GO:0016043 cellular component organization and biogenesis	413
GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	896
GO:0006259 DNA metabolic process (DNA metabolic process)	176
GO:0006350 transcription (transcription)	482
GO:0006464 protein modification process (protein modification process)	593
GO:0006412 translation (translation)	375
GO:0016265 death (death)	31
GO:0040007 growth (growth)	0
GO:0008152 metabolic process (metabolic process)	4028
GO:0009058 biosynthetic process (biosynthetic process)	824
GO:0005975 carbohydrate metabolic process	435
GO:0009056 catabolic process (catabolic process)	202
GO:0006091 generation of precursor metabolites and energy	389
GO:0006118 electron transport (electron transport)	290
GO:0006629 lipid metabolic process (lipid metabolic process)	169
GO:0015979 photosynthesis (photosynthesis)	73
GO:0019538 protein metabolic process (protein metabolic process)	1482
GO:0019748 secondary metabolic process (secondary metabolic process)	30
GO:0007275 multicellular organismal development (multicellular organismal development)	15
GO:0008938 abscission (abscission)	0
GO:0009790 embryonic development (embryonic development)	3
GO:0009791 post-embryonic development (post-embryonic development)	5
GO:0009908 flower development (flower development)	0
GO:0009835 ripening (ripening)	0
GO:0040029 regulation of gene expression, epigenetic	21
GO:0000003 reproduction (reproduction)	3
GO:0009856 pollination (pollination)	0
GO:0009628 response to abiotic stimulus (response to abiotic stimulus)	19
GO:0009607 response to biotic stimulus (response to biotic stimulus)	15
GO:0009719 response to endogenous stimulus (response to endogenous stimulus)	47
GO:0009605 response to external stimulus (response to external stimulus)	3
GO:0009606 tropism (tropism)	0
GO:0006950 response to stress (response to stress)	110
GO:0006810 transport (transport)	686
GO:0005575 cellular_component (cellular_component)	2487
GO:0005623 cell (cell)	2391
GO:0030312 external encapsulating structure	39
GO:0005618 cell wall (cell wall)	34
GO:0005622 intracellular (intracellular)	1663
GO:0005737 cytoplasm (cytoplasm)	676
GO:0005829 cytosol (cytosol)	16
GO:0005783 endoplasmic reticulum	91
GO:0005768 endosome (endosome)	0
GO:0005794 Golgi apparatus	28
GO:0005739 mitochondrion	65
GO:0005777 peroxisome (peroxisome)	7
GO:0009536 plastid (plastid)	24
GO:0005840 ribosome (ribosome)	282
GO:0005773 vacuole (vacuole)	4
GO:0005764 lysosome	1
GO:0005856 cytoskeleton (cytoskeleton)	66
GO:0005634 nucleus (nucleus)	549
GO:0005635 nuclear envelope	8
GO:0005730 nucleolus (nucleolus)	9
GO:0005654 nucleoplasm (nucleoplasm)	17
GO:0009579 thylakoid (thylakoid)	48
GO:0016020 membrane (membrane)	885
GO:0005886 plasma membrane (plasma membrane)	18
GO:0005576 extracellular region (extracellular region)	101
GO:0005615 extracellular space (extracellular space)	4
GO:0005578 proteinaceous extracellular matrix	3
GO:0005941 unlocalized protein complex (unlocalized protein complex)	20
GO:0003674 molecular_function (molecular_function)	6025
GO:0005488 binding (binding)	3674
GO:0030246 carbohydrate binding (carbohydrate binding)	60
GO:0003682 chromatin binding (chromatin binding)	5
GO:0008289 lipid binding (lipid binding)	45
GO:0003676 nucleic acid binding (nucleic acid binding)	1122
GO:0003677 DNA binding (DNA binding)	561
GO:0003700 transcription factor activity	174
GO:0003723 RNA binding (RNA binding)	150
GO:0008135 translation factor activity, nucleic acid binding	39
GO:0000166 nucleotide binding (nucleotide binding)	1223
GO:0019825 oxygen binding (oxygen binding)	2
GO:0005515 protein binding (protein binding)	700
GO:0005102 receptor binding (receptor binding)	36
GO:0003824 catalytic activity (catalytic activity)	3464
GO:0016787 hydrolase activity (hydrolase activity)	1208
GO:0004518 nuclease activity (nuclease activity)	52
GO:0016740 transferase activity (transferase activity)	1066
GO:0016301 kinase activity (kinase activity)	531
GO:0030234 enzyme regulator activity (enzyme regulator activity)	127
GO:0003774 motor activity (motor activity)	29
GO:0004871 signal transducer activity (signal transducer activity)	92
GO:0004872 receptor activity (receptor activity)	43
GO:0005198 structural molecule activity (structural molecule activity)	334
GO:0030528 transcription regulator activity (transcription regulator activity)	276
GO:0045182 translation regulator activity (translation regulator activity)	39
GO:0005215 transporter activity (transporter activity)	424

**Tableau 1.1 : Répartition des 15k gènes différentiellement exprimés en fonction de leur annotation en utilisant "The Gene Ontology" (GO). Le nombre de gènes considérés comme appartenant à chaque classe est indiqué en face de la classe correspondante.**

### 3. ANALYSE DU TRANSCRIPTOME LORS DU DEVELOPPEMENT DE LA GRAINE

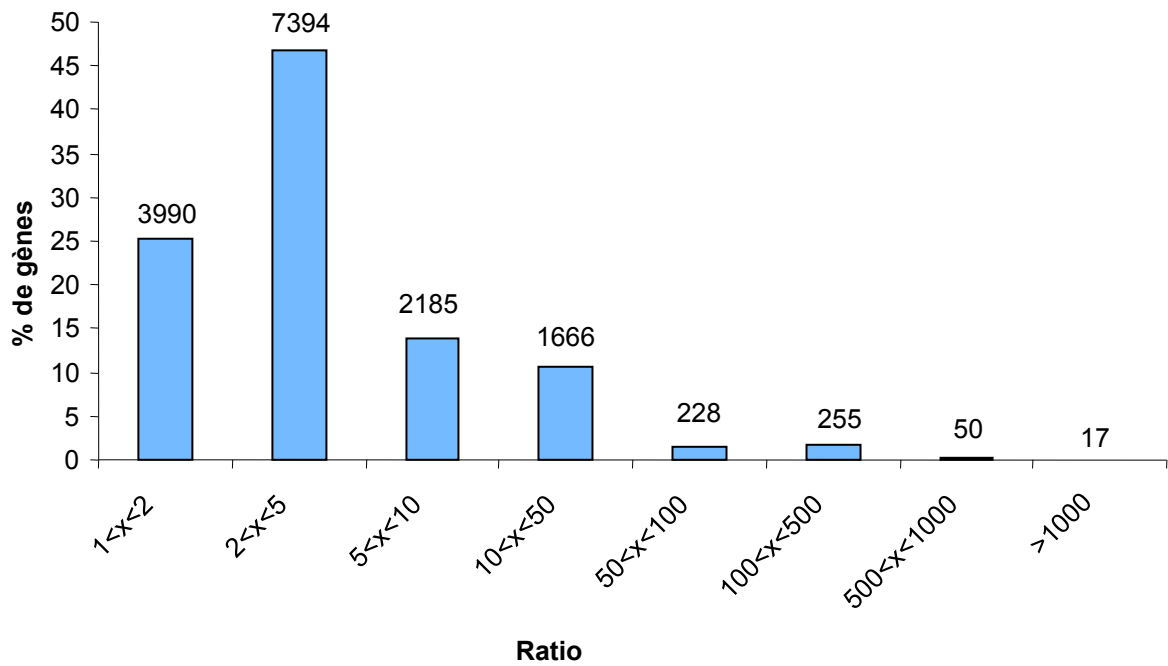
#### **3.1 Stratégie utilisée pour caractériser le développement de la graine**

Le principal objectif de cette étude est de décrire en détail les variations de profils d'expression des gènes pendant le développement de la graine et plus particulièrement pendant la phase de remplissage. Même si le changement dans le niveau d'expression d'un ARN messenger ne reflète pas toujours la quantité de protéines traduites et l'activité des enzymes, la présence de transcrits représente, cependant, une indication des mécanismes moléculaires mis en jeu. Par ailleurs, la puce Affymetrix récemment disponible chez *M. truncatula* ne dispose pas du génome entier mais elle contient approximativement 51000 gènes (50899 sondes) (approximativement 90% des gènes de *M. truncatula*), elle fournit donc l'opportunité unique de suivre l'expression d'un peu moins de 51k gènes tout au long du développement de la graine et ainsi permet d'avoir une vue d'ensemble des processus moléculaire de la graine chez cette espèce modèle.

Le profil d'expression de ces 51k gènes a été réalisé à six stades clefs du développement de la graine, stades choisis sur la base de données protéomiques (Gallardo *et al.* 2003). Tout d'abord, le stade 10 jours après pollinisation (JAP) correspond au stade d'embryogenèse tardive. Le stade 12 JAP marque la transition entre l'embryogenèse et le remplissage, caractérisée par la fin des divisions cellulaires de l'embryogenèse et précédant la synthèse des composés de stockage. Ensuite, les stades 16, 20 et 24 JAP correspondent au processus de remplissage du grain et de la synthèse des protéines de réserve, principalement composées de globulines 7S (vicilines et convicilines) ou 11S (légumines). Enfin, le stade 36 JAP correspond au stade de maturité physiologique de la graine et à l'entrée dans la phase de dessiccation.

#### **3.2 Analyse des profils d'expression pendant le développement de la graine**

En utilisant l'ADNc synthétisé à partir des stades précédemment cités, des hybridations ont été réalisées sur la puce "Affymetrix *M. truncatula* genome array". Des analyses statistiques ont été faites à partir de trois répétitions biologiques afin de mettre en évidence les gènes ayant des profils de niveau d'expression répétables. Un total de 92% (46811/50899) des sondes oligonucléotidiques présentes sur la puce ont pu s'hybrider avec de l'ADNc issu d'au moins un stade du développement de la graine. Par ailleurs, plus de la moitié de ces gènes (25745/46811) ont des transcrits détectables aux six stades étudiés. Ce premier résultat souligne que le développement de la graine est un processus complexe qui implique la grande majorité des gènes. Une analyse statistique de type ANOVA a permis d'identifier 15786 gènes dont l'expression varie significativement durant le développement de la graine (p-value <0,01). Une première structuration du jeu de données a été réalisée en annotant ces 15k gènes en fonction de leur domaine protéique (annotation utilisant The Gene Ontology Project, <http://www.geneontology.org/index.shtml>) (tableau 1.1). Un histogramme catégorisé reflétant les amplitudes de variations d'expression des gènes (figure 1.1) a révélé que, parmi les 15k gènes différenciellement exprimés, pour approximativement 75% d'entre eux (11795/15786) cette expression varie de plus de deux fois entre les différents stades du développement de la graine. Cette variation d'expression atteint même un facteur 500 pour



**Figure 1.1 :** Histogramme catégorisé reflétant les amplitudes de niveaux d'expression des 15k gènes au cours du développement de la graine. Les ratios d'expression sont indiqués en abscisse et le pourcentage de gènes concernés en ordonnée. Le chiffre présent au dessus des histogrammes correspond au nombre de gènes concernés par cette variation.

67 gènes. Parmi ces derniers, 16 (16/67) correspondent à des gènes codant pour des protéines de maturation de la graine (plusieurs protéines de réserves et des protéines associées à la dessiccation de type Late Embryogenesis Abundant, les protéines LEA).

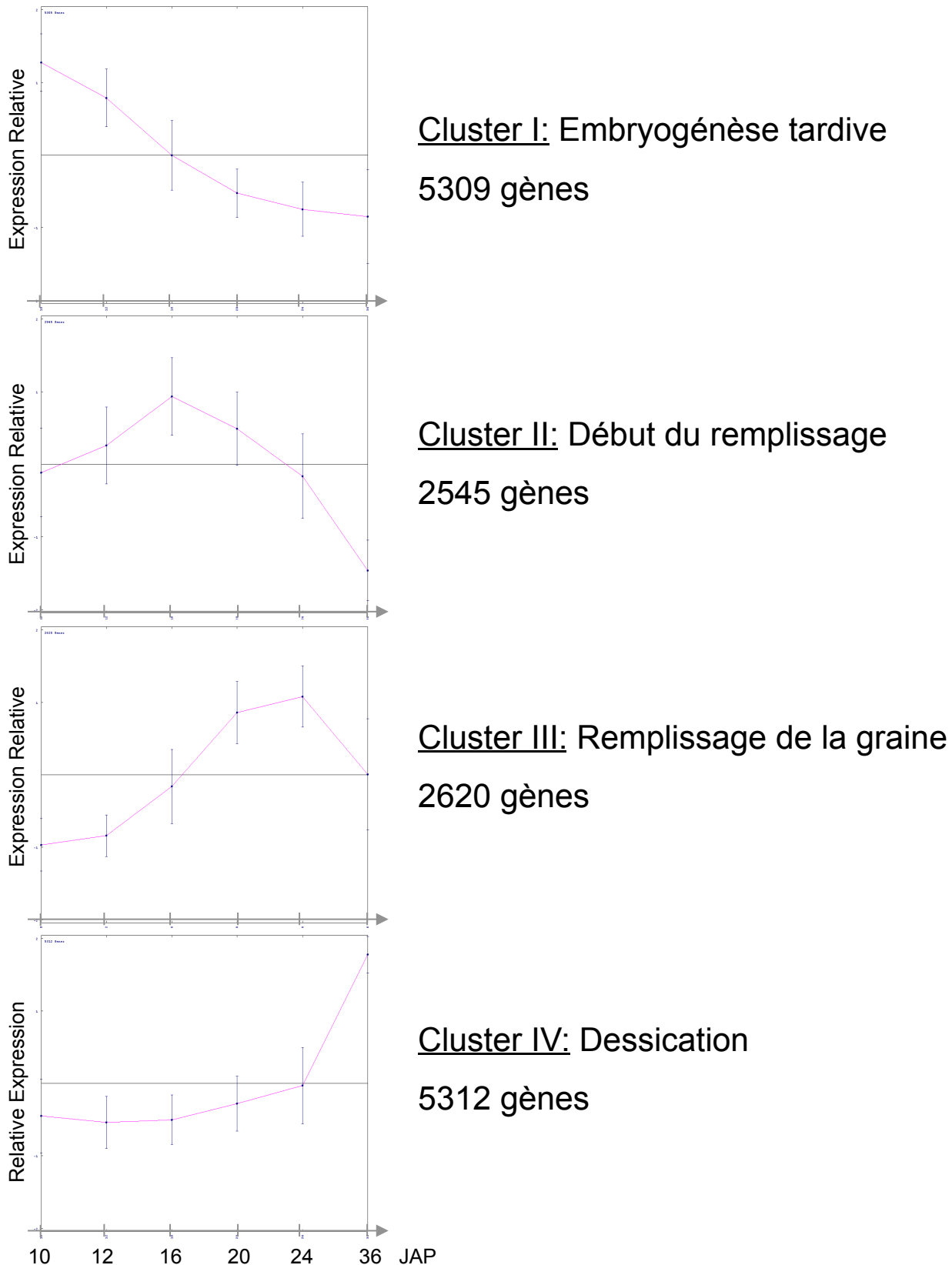
Un classement hiérarchique, sur la base des profils d'expression, des 15k gènes différentiellement exprimés a permis de mettre en évidence quatre groupes (Figure 1.2). Le premier groupe (cluster I) contient 5309 gènes qui présentent un maximum d'expression à 10 JAP, donc associés à des mécanismes d'embryogenèse. Le deuxième et le troisième groupe (cluster II et III) sont constitués de gènes qui sont liés aux mécanismes de remplissage du grain. Le cluster II comporte 2545 gènes préférentiellement exprimés à 16 JAP, ce groupe contient des gènes associés au début de la synthèse des ARNm des protéines de réserve : 14 JAP pour la synthèse des ARNm de vicilines et 16 JAP pour les légumineuses (Verdier *et al.* 2008). Le cluster III est composé de 2620 gènes qui présentent un pic d'expression à 20 et/ou 24 JAP, c'est à dire de manière concomitante au maximum d'expression des protéines de réserve (20 JAP pour les vicilines et 24 JAP pour les légumineuses K) (Verdier *et al.* 2008). Enfin, le cluster IV est composé de 5312 gènes dont le maximum d'expression se situe à 36 JAP, ce qui correspond à la phase de dessiccation.

### **3.3 Classification fonctionnelle des gènes différentiellement exprimés**

Afin d'identifier les principaux processus moléculaires associés aux différents stades du développement de la graine, nous avons utilisé l'outil GeneBins (Goffard *et al.* 2006 et 2007) pour assigner à chaque transcrite une classe fonctionnelle selon l'ontologie KEGG (the Kyoto Encyclopedia of Genes and Genomes ontology). Environ 17,5% (2767/15786) des 15k gènes différentiellement exprimés ont pu être associés à une classe fonctionnelle (p-value <0,05 avec correction de Bonferroni). En accord avec les besoins en azote et carbone de la graine, on peut noter que les classes fonctionnelles les plus représentées sont le métabolisme des carbohydrates et celui des acides aminés. La répartition des gènes dans les classes fonctionnelles a été réalisée pour les différents clusters d'expression. La première conclusion est que selon le cluster d'expression et donc en fonction du stade de développement de la graine, la répartition des classes fonctionnelles varie fortement (Figure 1.3), traduisant des fonctions biologiques différentes aux stades étudiés.

L'ontologie KEGG est organisée en six différents modules dont quatre sont applicables aux plantes : les voies métaboliques (metabolic pathways), les mécanismes de traitement de l'information génétique (genetic information processing), les mécanismes de traitement de l'information environnementale (environmental information processing) et les mécanismes cellulaires (cellular processes) (Kanehisa *et al.* 2004).

Nous pouvons ainsi montrer que les différents modules sont préférentiellement associés avec des clusters spécifiques. En effet, nous constatons une grande abondance de transcrits associés aux différentes activités métaboliques pendant le remplissage du grain (cluster II et III). A l'opposé, au cours de l'embryogenèse (cluster I) et de la dessiccation (cluster IV), les transcrits sont plutôt associés aux mécanismes de traitement de l'information génétique. Enfin, les mécanismes de traitement de l'information environnementale ainsi que les



**Figure 1.2 :** Représentation des quatre clusters hiérarchiques des 15k gènes différentiellement exprimés durant le développement de la graine. Le nom des clusters, ainsi que le processus auquel il est rattaché et le nombre de gènes qui le compose sont indiqués. Les graphiques représentent pour chaque cluster les moyennes des expressions relatives de tous les gènes présents dans le cluster (+/- écart-types).

mécanismes cellulaires sont majoritairement associés avec des gènes présents respectivement dans le cluster III (expression à 20-24 JAP) et dans le cluster I (expression à 10 JAP).

Pour approfondir les mécanismes impliqués tout au long du développement de la graine, nous avons disséqué ce jeu de données.

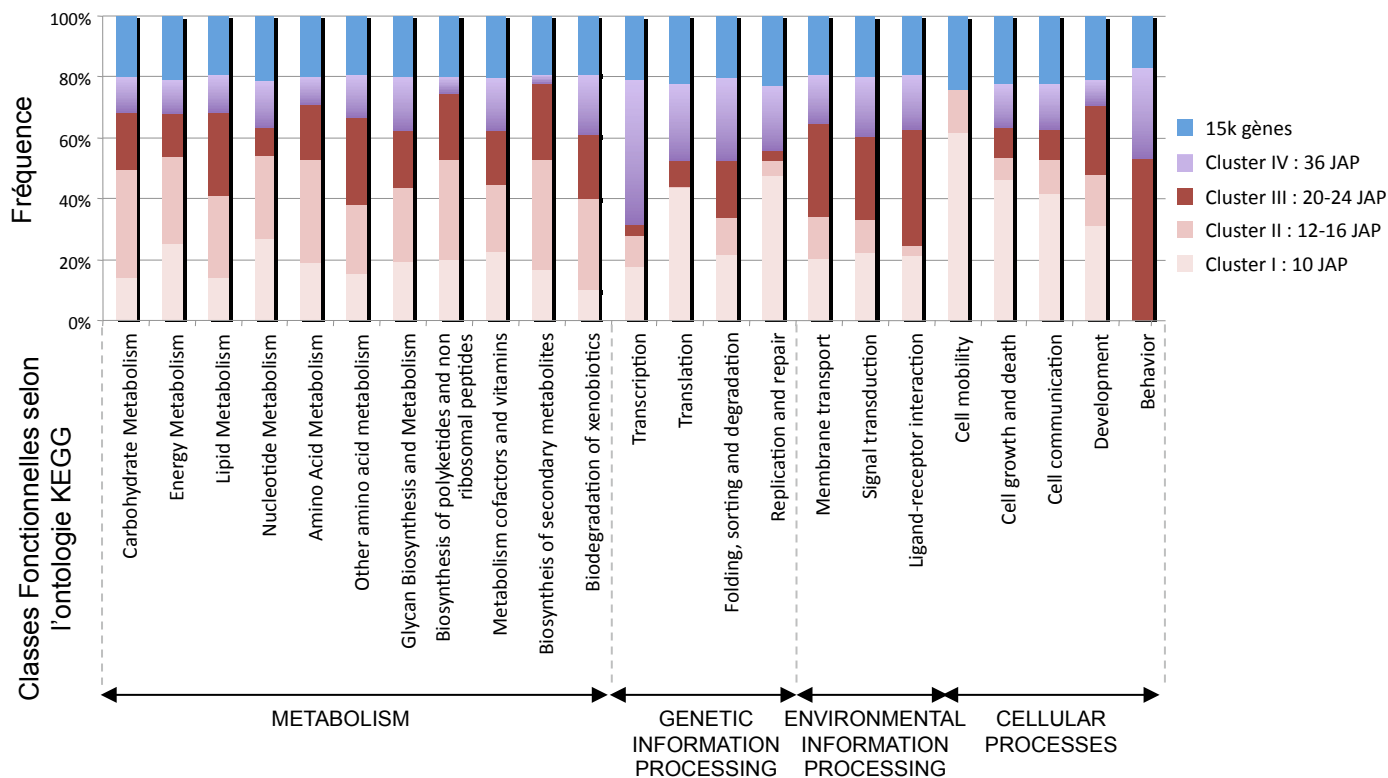
### **3.4 Processus associés aux différentes étapes du développement de la graine chez *M. truncatula***

#### **3.3.1 L'embryogenèse**

La phase d'embryogenèse est définie par des étapes successives d'histodifférentiation de l'embryon : stade globulaire, cœur, torpille et finalement cotylédonaire. Durant cette période, l'embryon acquiert sa polarité selon l'axe apical-basal avec la mise en place des méristèmes. Dans notre étude, nous avons identifié un groupe de gènes exprimés à 10 JAP pouvant intervenir dans ces processus. Ce groupe est composé de gènes appartenant à plusieurs classes fonctionnelles dont principalement celles associées à la mobilité cellulaire, la croissance cellulaire, la réplication/réparation de l'ADN et la communication cellulaire. Tous ces processus sont des acteurs majeurs de la division cellulaire et sont le reflet d'une intense activité mitotique.

Cette activité importante de division cellulaire intense est cohérente avec la physiologie de l'embryon qui, au stade 10 JAP, est en fin de morphogenèse. En effet, l'embryon est au début du stade torpille et n'occupe qu'une faible proportion du volume de la graine. Cette intense division cellulaire se prolonge jusqu'au stade 12 JAP lorsque l'expansion cellulaire devient le processus majeur de croissance. Parallèlement, à 10 JAP, l'albumen est aussi soumis à une active prolifération cellulaire associée à une forte activité d'endoploïdie. Les cellules de l'albumen contiennent de fortes quantités de mitochondrie indiquant une forte activité pour soutenir le développement de l'embryon et le remplissage du grain en fournissant les métabolites essentiels (pour revue, Seed biology of *M. truncatula*, 2006, <http://www.noble.org/MedicagoHandbook/>). De plus, notons que c'est à la fin de ce stade de division cellulaire intense que le nombre de cellules de l'embryon est déterminé. Une relation a été établie chez le pois entre le taux de croissance de la graine et le nombre de cellules des cotylédons (Munier-Jolain *et al.* 1998). C'est donc à ce stade de division cellulaire précédant le remplissage que la graine acquiert son potentiel de croissance. En accord avec ces résultats, les gènes présents dans ce cluster sont des candidats potentiels impliqués dans l'établissement de ce potentiel de croissance et donc dans l'établissement de la taille finale de la graine.

Les gènes de la mobilité cellulaire sont impliqués dans la régulation du cytosquelette d'actine jouant ainsi un rôle dans le maintien de la structure cellulaire. Les gènes de croissance cellulaire sont des gènes impliqués dans le cycle cellulaire lui-même ou dans le processus d'apoptose qui régule la prolifération cellulaire. Les gènes de réplication/réparation identifiés sont essentiellement des ADN polymérases, enzymes qui permettent la réplication de l'ADN en catalysant la polymérisation des désoxyribonucléotides le long du brin d'ADN. Enfin, les mécanismes de communication cellulaire permettent de contrôler les activités cellulaires de base et de



**Figure 1.3 :** Classification fonctionnelle des 15k gènes différentiellement exprimés durant le développement de la graine selon l'ontologie KEGG en fonction de leur répartition dans les différents clusters (Clusters I à IV). La répartition fonctionnelle de l'ensemble des gènes, sans tenir compte des clusters, est également représentée sous le nom de groupe "15k gènes" (en bleu).

coordonner ces processus cellulaires. Nous constatons également que les enzymes de biosynthèse des nucléotides sont aussi exprimés entre 10 et 16 JAP pour fournir les nucléotides aux processus de réplication ainsi qu'au métabolisme énergétique par l'intermédiaire de l'ATP, molécule essentielle à l'activité de beaucoup d'enzymes métaboliques.

### 3.3.2 Le remplissage du grain

Durant la phase de remplissage, le poids de la graine augmente et des macromolécules de réserve (protéines, lipides et carbohydrates) s'accumulent. Dans notre étude, nous avons mis en évidence deux groupes distincts de gènes préférentiellement exprimés durant cette période : le cluster II constitué de gènes avec un maximum d'expression à 16 JAP, au début de la phase de remplissage, et le cluster III composé de gènes exprimés plus tardivement à 20-24 JAP. Même si ces deux groupes correspondent au même processus développemental, nous pouvons constater des différences dans les processus métaboliques les caractérisant. Pour analyser quels sont les processus clefs durant ces étapes de remplissage, nous avons utilisé l'outil PathExpress (disponible sur internet : <http://bioinfoserver.rsbs.anu.edu.au/utis/PathExpress/>) qui permet de mettre en évidence les voies métaboliques qui sont les plus soumises à régulation à partir de données d'expression (Goffard *et al.* 2007). Le résultat de cette analyse nous indique que cinq voies métaboliques (p-value <0,1 ajustée avec le facteur correctif False Discovery Rate, FDR) sont préférentiellement associées au cluster II (16 JAP). Ce sont les voies métaboliques des carbohydrates, des acides aminés, des lipides, et des métabolites secondaires.

Le métabolisme des carbohydrates représente dans notre étude le processus métabolique le plus fortement régulé à ce stade de développement. Ce sont en particulier les voies de la glycolyse/néoglucogenèse et celles associées à la biosynthèse/dégradation de l'amidon et du saccharose. Ce résultat est cohérent avec le rôle central du métabolisme du sucre durant le développement de la graine, principalement au début du remplissage. En effet, même si le contenu de la graine mature de *M. truncatula* n'est composé que respectivement de 6-10 % de sucres solubles et moins de 1 % d'amidon (Djemel *et al.* 2005), ces molécules permettent le maintien d'une source importante de carbone durant le développement. De plus, le saccharose possède un double rôle : il est à la fois une molécule centrale dans le métabolisme du sucre et une molécule « signal ». En effet, ce rôle de signalisation est joué par la balance saccharose/hexose qui représente un complexe de régulation de la synthèse des molécules de réserve par l'intermédiaire des saccharoses synthases et des invertases (pour revue Koch 2004 ; Weber *et al.* 1997 ; Smeekens *et al.* 2000).

La forte régulation des ARNm de gènes impliqués dans le métabolisme des acides aminés, est cohérente avec une forte synthèse d'acides aminés à ce stade, essentiels pour assurer la synthèse de protéines de stockage. En effet, ces protéines, essentiellement accumulées dans l'embryon, représentent 32-34% du poids sec de la graine de *M. truncatula* (Djemel *et al.* 2005). Elles appartiennent majoritairement à la famille des globulines et sont de trois types : les légumine, les vicilines et les convicilines. Ces protéines ont la particularité de s'accumuler de façon séquentielle, contrairement à *A. thaliana*, avec dans un premier temps la synthèse des vicilines puis plus





tardivement la synthèse des convicilines et légumines (Gatehouse *et al.* 1982 ; Gatehouse *et al.* 1986 ; Gallardo *et al.* 2003 ; Verdier *et al.* 2008).

Dans ce même cluster, on constate que les gènes du métabolisme des lipides sont sur-représentés, en particulier ceux associés à la biosynthèse des acides gras. Comme pour le métabolisme des acides aminés, cette observation est cohérente avec le début de la phase de remplissage où la graine synthétise des acides gras qui seront incorporés dans des triacylglycérides, principalement sous forme C16 :0, C18 :1, C18 :2 et C18 :3 qui représenteront 95% des acides gras totaux de la graine de *M. truncatula* (Djemel *et al.* 2005). A maturité, 8 à 10% de son poids sec est composé d'acides gras de réserve.

Enfin, la dernière voie métabolique très représentée à ce stade est celle du métabolisme secondaire. La forte implication du métabolisme secondaire au début du remplissage avec en particulier la voie de biosynthèse des flavonoïdes, nous indique que c'est également durant ce stade que la graine met en place des mécanismes de pigmentation. De nombreuses enzymes associées à la biosynthèse des flavonoïdes peuvent être identifiées. Ces enzymes de biosynthèse permettent l'accumulation d'isoflavones dans l'embryon (Dhaubhadel *et al.* 2003) ou d'anthocyanes et de flavonols nécessaires à la pigmentation des téguments (Lepiniec *et al.* 2006).

Le second groupe de gènes associé au remplissage de la graine est le cluster III (20-24 JAP). Les gènes de ce cluster sont globalement plus associés à diverses voies métaboliques. De plus, on constate également qu'un certain nombre de gènes impliqués dans le traitement de l'information environnementale sont fortement représentés dans ce cluster. Ces gènes sont souvent associés à des transporteurs et à des kinases indiquant une intense signalisation cellulaire. Cette importante communication cellulaire pourrait intervenir dans les processus de fin de remplissage et de préparation à la dessiccation.

Notons qu'en considérant la période complète de remplissage (clusters II et III), le traitement de l'information génétique est la classe fonctionnelle la plus sous représentée. Cette faible représentativité des mécanismes de réplication, transcription et traduction durant la phase de remplissage peut indiquer que le déclenchement du programme génétique aboutissant au remplissage est principalement régulé lors du stade de transition entre embryogenèse et remplissage, et que par conséquent c'est un stade clef dans la régulation de cette phase. Ce résultat est en accord avec l'identification d'un pool important de protéines ribosomales à 12 JAP dans les noyaux de graines de *M. truncatula* (Repetto *et al.* 2008)

### 3.3.3 La dessiccation

La dernière phase de développement de la graine, avant la germination, est la préparation à la dessiccation. Cette étape est caractérisée par une forte et rapide perte d'eau. Le cluster IV, qui est composé de gènes dont le pic d'expression à 36 JAP est associé avec cette étape de dessiccation. Nos résultats suggèrent que ce stade présente une activité métabolique réduite : en effet, peu d'ARNm codants pour les enzymes du métabolisme sont représentés.



Ce résultat est en accord avec l'entrée de la graine en quiescence métabolique. En effet, on considère qu'à 36 JAP la graine atteint un stade de maturité physiologique. Par ailleurs, notre analyse révèle que les transcrits associés aux processus transcriptionnels s'accumulent, incluant les ARN polymérase et les facteurs de transcription. Cette intense activité transcriptionnelle reflète la préparation d'un pool d'ARNm au stade 36 JAP. En effet, des travaux ont montré qu'un pool d'ARNm était stocké durant le développement de la graine chez *A. thaliana*, pool servant lors de la germination pour synthétiser rapidement des protéines (Rajjou *et al.* 2003).



## 4. CONCLUSION DU CHAPITRE

### Synthèse des résultats

Avant ce travail, des analyses de suivi d'expression de gènes avaient déjà été réalisées sur la légumineuse modèle, *M. truncatula*, dans différents laboratoires et sur différents tissus. Cependant, aucune uniformisation des méthodes et des données n'avaient été réalisées. En adoptant chez *M. truncatula* une démarche comparable à celle entreprise sur d'autres espèces modèles, Ce travail a permis de **générer une base de données d'expression pour 51k gènes chez *M. truncatula* durant le développement des graines**. Combinée aux études d'expression dans les autres tissus (feuilles, pétioles, fleurs, tiges, bourgeons végétatifs, gousses et racines), cette étude offre l'opportunité d'identifier des gènes spécifiques de la graine, souvent associés à des processus spécifiques. Une comparaison des mécanismes de régulation de l'expression des gènes avec d'autres espèces modèles peut maintenant être envisagée afin de déterminer la conservation ou non des mécanismes moléculaires intervenant dans les différents processus développementaux.

L'analyse d'expression des gènes à l'intérieur d'organes en cours de développement a permis **d'identifier les processus physiologiques** importants et ainsi mettre en évidence **des stades clefs du développement**. Nous avons réalisé ce type d'analyse pour le développement de la graine. Quatre phases clefs ont été identifiées, et leurs processus physiologiques disséqués.

Tout d'abord, la phase associée à l'embryogenèse tardive (10 JAP) qui est marquée par une intense activité mitotique qui est déterminante dans la détermination de la taille finale du grain. Ensuite, le remplissage du grain qui est composé de deux phases : le début du remplissage marqué par une forte activité métabolique et la fin du remplissage présentant également une importante activité métabolique couplée à des processus de signalisation cellulaire. Enfin, la dessiccation qui présente une très forte réduction de l'activité métabolique et une augmentation des processus transcriptionnels. Par ailleurs, une faible activité transcriptionnelle est constatée durant les phases de remplissage à 16 JAP et 20-24 JAP, ce qui pourrait signifier que le déclenchement et la régulation du remplissage des graines se produit lors de la phase de transition entre la fin de l'embryogenèse et le début de la synthèse des protéines de réserve, donc entre les stades allant de 12 à 16 JAP.

### Perspectives et applications

Cette **base de données, disponible à toute la communauté scientifique**, est accessible sur le site internet : <http://bioinfo.noble.org/gene-atlas/> et constitue une ressource essentielle pour les laboratoires travaillant sur les légumineuses.

Dans le futur, de nouvelles analyses réalisées avec cette même puce de gènes mais dans d'autres conditions (stress, effet de mutation, ...) pourraient servir à compléter cette base de données et ainsi la rendre plus exhaustive.



**CHAPITRE 2 : Identification de  
facteurs de transcription impliqués  
dans le remplissage de la graine de  
*M. truncatula* et isolement de  
mutants pour ces gènes**





## CHAPITRE 2

### Identification de facteurs de transcription candidats impliqués dans le remplissage de la graine de *M. truncatula* et isolement de mutants pour ces gènes

#### 1. OBJECTIFS ET DEMARCHE

De nombreux travaux ont permis de mettre en évidence des facteurs de transcription (TFs) essentiels pour le développement de la graine chez les dicotylédones. Tout d'abord, les analyses de mutants affectés au niveau des graines matures ont permis l'identification de nombreux gènes clefs associés au développement tels que *LEC1*, *LEC2*, *FUS3* ou *ABI3* (Meinke *et al.*, 1994 ; Parcy *et al.*, 1997 ; Stone *et al.*, 2001). Puis avec l'émergence de la génomique et des approches à haut-débit, les premières études globales ont permis d'identifier des gènes candidats pouvant être impliqués dans les différentes étapes du développement de la graine (Girke *et al.* 2000 ; Ruuska *et al.* 2002). Ces approches utilisant des puces à ADN ont été réalisées sur *A. thaliana* et centrées sur le remplissage et la maturation des graines. Plus récemment, de Folter *et al.* (2004) ont focalisé leur étude sur l'expression des facteurs de transcription dans la graine d'*A. thaliana* par hybridation sur filtre haute densité.

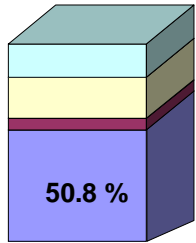
Au début de ce travail de thèse, les données transcriptomiques dont nous disposions sur le développement de la graine des légumineuses se résumaient à l'analyse faite par Gallardo *et al.* (2007), analyse réalisée avec la puce Mt16kOLI1 élaborée dans le cadre du projet GLIP par le laboratoire d'Helge Küster (Bielefeld, Germany). Les résultats de cette étude, bien que robustes, ne permettaient pas d'obtenir de données complètes sur la régulation des TFs en raison d'abord de la sensibilité de la méthode et ensuite de la faible représentation des TFs dans les banques d'EST et donc par conséquent sur la puce.

Suite au manque de données d'expression disponibles sur la régulation transcriptionnelle de type gène spécifique lors du développement de la graine chez les légumineuses, nous nous sommes donc proposés d'élargir les connaissances sur ce sujet. L'objectif de ce travail a donc été :

- (i) de générer une base de données d'expression des facteurs de transcription lors du développement de la graine de *M. truncatula*,
- (ii) de comparer la régulation transcriptionnelle chez *M. truncatula* à ce que l'on connaît chez les autres dicotylédones, et principalement *A. thaliana*,

(a)

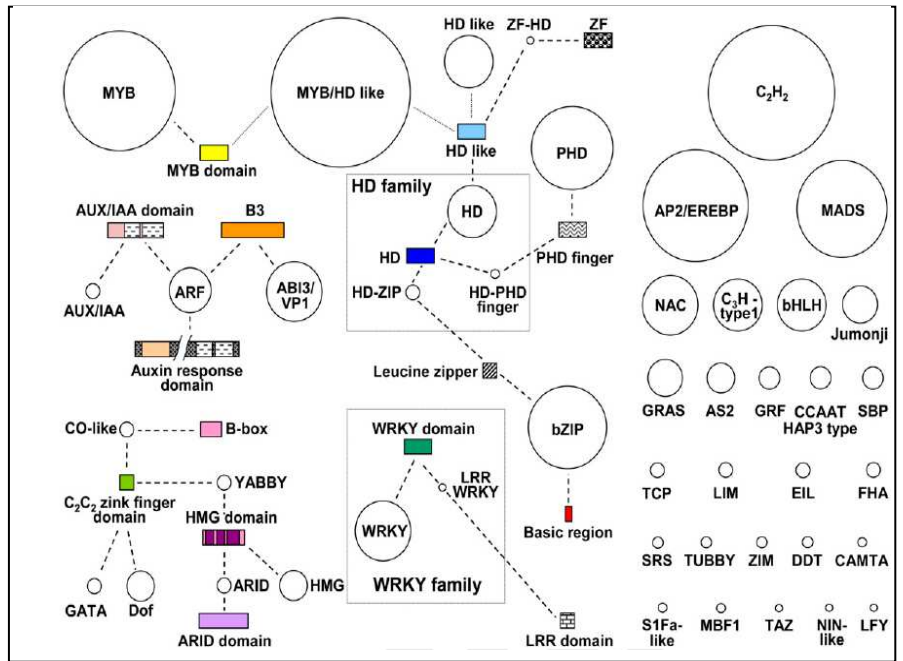
~1400 putatifs TFs :



712 TFs

- Autres
- Transduction du signal
- Remodelage de la chromatine
- Facteurs de transcription

(b)



(Figure extraite de Udvardi et al., 2007)

**Figure 2.1 : Collection de TFs disponible sur la plateforme de qRT-PCR**

(a) Représentation des 1400 TFs de la collection de départ.

(b) Représentation des familles de la collection des 712 « vrais » TFs disponibles. Les familles sont représentées par des cercles dont la taille dépend du nombre de TFs qui la compose. Les relations entre les différentes familles sont également indiquées. Cette figure est basée sur la figure de Riechmann (2001) et est extraite de Udvardi *et al.*(2007).

- (iii) de fournir une liste de candidats "expressionnels" pouvant être des acteurs clefs du développement de la graine et particulièrement de l'étape de maturation,
- (iv) d'isoler des mutants pour les différents facteurs de transcription candidats.

Pour y parvenir, nous avons choisi d'utiliser la récente plateforme de PCR quantitative en temps réel disponible au Max Planck Institute à Golm (Germany), coordonnée par Michaël Udvardi. L'espèce modèle, *M. truncatula*, a été choisie pour réaliser cette étude et sept stades de développement de la graine ont été étudiés. Les profils d'expression ont été réalisés sur plus de 1400 facteurs de transcription putatifs identifiés *in silico* grâce aux annotations automatiques du génome de *M. truncatula* par l'équipe de Christopher Town (The Institute for Genomic Research, Rockville, USA) (Udvardi *et al.* 2007). Ensuite, l'analyse approfondie des domaines protéiques conservés de ces gènes (INTERPRO, <http://www.ebi.ac.uk/interpro/>) a révélé que seulement 712 gènes correspondaient réellement à des facteurs de transcription (Figure 2.1). Le suivi d'expression a été réalisé sur la collection de plus de 1400 TFs putatifs, cependant l'analyse des résultats présentés dans la suite du chapitre est centrée sur les 712 "vrais" TFs.

Le chapitre est donc constitué dans un premier temps de l'analyse des résultats de cette étude, sous la forme d'une publication, présentant les profils d'expression de plus de 700 facteurs de transcription au cours du développement de la graine, ainsi que de la comparaison entre l'expression des TFs identifiés chez *M. truncatula* et ceux connus chez *A. thaliana*. Puis, une seconde partie présente les candidats "expressionnels" que nous avons sélectionné suite à cette étude. Et enfin, une troisième partie décrit les différents outils de génomique fonctionnelle disponibles chez *M. truncatula* et l'isolement de mutants pour les gènes sélectionnés.



## **2. LE CONTROLE TRANSCRIPTIONNEL DU DEVELOPPEMENT DE LA GRAINE *M. TRUNCATULA***

Ce travail se réfère à un article publié dans la revue Plant Molecular Biology intitulé :

### **Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling**

**Jérôme Verdier**, Klementina Kakar, Karine Gallardo, Christine Le Signor, Grégoire Aubert, Armin Schlereth, Christopher D. Town, Michael K. Udvardi and Richard D. Thompson

Plant Molecular Biology (2008) Vol. 67(6) : 567-580

#### Résumé :

Les graines des légumineuses constituent une source importante de protéines pour l'alimentation. La plante modèle *M. truncatula* est caractérisée par un processus de développement de la graine très proche de celui des autres légumineuses, impliquant l'interaction d'un ensemble de facteurs de transcription. Dans cette étude, nous reportons la première analyse de suivi d'expression de plus de 700 gènes de *M. truncatula* codant pour des facteurs de transcription putatifs, réalisée en utilisant la PCR quantitative en temps réel. 169 de ces TFs ont été identifiés comme étant exprimés durant l'embryogenèse tardive, le remplissage ou la déshydratation du grain. L'expression dans les différents tissus de la graine a été analysée pour les 41 TFs les plus fortement exprimés. Dans le but d'identifier les gènes cibles putatifs de ces TFs, ces résultats d'expression ont été combinés avec un jeu de données issu du transcriptome des graines. Cette étude a permis d'identifier 135 gènes co-exprimés avec 17 TFs tissus-spécifiques incluant des cibles potentielles. Certains de ces TFs co-exprimés avec des ARNm de protéines de réserve de *M. truncatula* correspondaient à ceux connus pour réguler la synthèse des protéines de réserve chez *A. thaliana*, alors que les profils d'expression d'autres TFs pourraient être spécifiquement associés au décalage d'expression des légumineuses, décalage seulement observé chez les légumineuses.



# Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling

Jérôme Verdier · Klementina Kakar · Karine Gallardo · Christine Le Signor ·  
Grégoire Aubert · Armin Schlereth · Christopher D. Town · Michael K. Udvardi ·  
Richard D. Thompson

Received: 8 November 2007 / Accepted: 13 March 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Legume seeds represent a major source of proteins for human and livestock diets. The model legume *Medicago truncatula* is characterized by a process of seed development very similar to that of other legumes, involving the interplay of sets of transcription factors (TFs). Here, we report the first expression profiling of over 700 *M. truncatula* genes encoding putative TFs throughout seven stages of seed development, obtained using real-time quantitative RT-PCR. A total of 169 TFs were selected which were expressed at late embryogenesis, seed filling or desiccation. The site of expression within the seed was examined for 41 highly expressed transcription factors out of the 169. To identify possible target genes for these TFs, the data were combined with a microarray-derived transcriptome dataset. This study identified 17 TFs preferentially expressed in individual seed tissues and 135 corresponding co-expressed genes, including possible targets. Certain of the TFs co-

expressed with storage protein mRNAs correspond to those already known to regulate seed storage protein synthesis in Arabidopsis, whereas the timing of expression of others may be more specifically related to the delayed expression of the legumin-class storage proteins observed in legumes.

**Keywords** Legume seed development · Transcription factors · Storage protein · Embryo · Endosperm · Seed coat

## Introduction

More than 70% of the average human diet is derived from seeds (essentially cereals and legumes). Seeds of cereal species, such as maize (*Zea mays*) and wheat (*Triticum aestivum*) are a major source of starch and contain less than 16% protein content in their main storage tissue, the endosperm. Legume species such as pea (*Pisum sativum*), soybean (*Glycine max*) or fava bean (*Vicia faba*) represent a richer protein source than the major cereals, with seed protein content ranging from 20% to as much as 40%, depending on species, genotype and environment. In most legumes, the storage compounds accumulate essentially in the embryo which develops surrounded by endosperm and enclosed in the maternal integument layers. Legume seed development commences with embryogenesis during which the embryo undergoes a programme of cell divisions, acquires polarity along the shoot-root axis, and the shoot and root apical meristems are established. In the subsequent seed maturation or seed filling period, embryo cell divisions cease and cell expansion occurs. This stage corresponds to the deposition of storage molecules, lipids, carbohydrates and proteins. The economically and nutritionally important legume storage proteins consist principally of the legumin (11S) and vicilin (7S) globulin classes.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-008-9320-x) contains supplementary material, which is available to authorized users.

J. Verdier · K. Gallardo · C. Le Signor · G. Aubert ·  
R. D. Thompson (✉)  
Unité Mixte de Recherche en Génétique et Ecophysiologie des  
Légumineuses à Graines (UMR-LEG), Institut National de la  
Recherche Agronomique (INRA), Domaine d'Epoisses, 21110  
Bretenieres, France  
e-mail: thompson@epoisses.inra.fr

K. Kakar · A. Schlereth · M. K. Udvardi  
Institute for Molecular Plant Physiology, 14476 Golm, Germany

C. D. Town  
The Institute for Genomic Research (TIGR), Rockville, MD  
20850, USA

M. K. Udvardi  
Samuel Robert Noble Foundation, Ardmore, OK 73401, USA





The extent of transcriptional regulation in the developing *M. truncatula* seed has been evaluated by a comparison between transcriptome and proteome profiles during seed development. This suggested that the abundances of approximately 50% of the proteins identified, including the seed storage proteins, were mainly regulated at the transcriptional level (Gallardo et al. 2007). In plants, gene regulation at the level of transcription is fundamental to many aspects of development. Transcription factors (TFs) account for nearly 6% of the gene complement of the Arabidopsis genome sequence, with many of them being plant-specific (Riechman and Ratcliffe 2000).

Profiling of TF gene expression during seed development has been carried out for Arabidopsis and cereals (Ruuska et al. 2002; De Folter et al. 2004; Druka et al. 2006), and several TFs have been independently identified by mutant analyses as playing key roles in different steps of seed development. Examples include regulators of the maternal contribution to seed size (*TTG2*, Garcia et al. 2005), homeotic mutants of embryo development (*MONOPTEROS*, Hardtke and Berleth 1998), factors governing the deposition of seed storage proteins (*OPAQUE2*, Unger et al. 2002; *LEAFY COTYLEDON2*, Santos Mendosa et al. 2005), and lipids (*WRINKLED1*, Cernac and Benning 2004), and those directing other aspects of seed maturation (e.g., desiccation tolerance, *ABSCISIC ACID-INSENSITIVE3*, Parcy et al. 1995).

The aim of the present study was to provide an overview of transcription factor genes expressed during legume seed development. We have chosen to use the model legume *Medicago truncatula* for which extensive sequence information is available from both EST and genome sequencing programmes (Cook 1999; Ané et al. 2008). The strategy we employed involved a large-scale profiling of transcription factor gene expression during seed development, using quantitative real-time RT-PCR. Temporal changes in gene expression of more than 700 *M. truncatula* TF transcripts were identified and catalogued during seed development. A combination of our data with a microarray transcriptome dataset (Gallardo et al. 2007) has allowed us to assign TFs to key phases and processes during seed development. About 80 TF genes were preferentially expressed during storage compound accumulation, most of which could be attributed to one of the three principal seed tissues. We have further identified amongst these TF candidates for regulating the sequential expression of vicilin and legumin storage protein classes observed in legumes.

## Materials and methods

### Plant growth

Two independent series of *M. truncatula* plants (cultivar Jemalong, line A17) representing two biological replicates

were grown and tagged according to Gallardo et al. (2007). ~5000 seeds [1.5 g of seeds for each of the seven developmental stages ranging from 10 to 36 Days After Pollination (DAP)] were collected, on ice to prevent dehydration, per biological replicate. For seed tissue analysis, seed coat, endosperm and embryo were dissected from 14 DAP seeds under a magnifying glass (magnification,  $\times 3.5$ ) on Petri dishes placed on ice and immediately frozen in liquid nitrogen after tissue dissection. Flowers, leaves, stems, and roots were collected in growth chamber at flowering.

### Total RNA isolation and cDNA synthesis

Total RNA was isolated from developing seeds according to Chang et al. (1993). 200  $\mu\text{g}$  of total RNA was treated with RNase free DNase RQ I<sup>TM</sup> (Promega, Madison, USA) for 45 min according to the manufacturer's instructions. The absence of genomic DNA contamination was confirmed by real-time RT-PCR on non-reverse-transcribed RNA. RNA was purified and concentrated using a centrifugal filter (Microcon YM-30, Millipore, Bedford, USA) before use in first strand cDNA synthesis. RT reactions were performed using a mix of polyT(23) and random primers with SuperScript<sup>TM</sup> III reverse transcriptase (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions. The efficiency of cDNA synthesis was assessed by real-time PCR amplification of a control gene encoding *Ubiquitin*. Only cDNA preparations, that yielded similar threshold cycle (Ct) values (i.e.  $24 \pm 1$ ) and that showed a 3'/5' ratio in the range of 1–1.5 (using formula  $2^{(\text{Ct}_{5'} - \text{Ct}_{3'})}$ ) for the *Ubiquitin* gene, were used for comparing TF transcript levels. cDNA from the three seed tissues was synthesised using iScript one-step reverse-transcriptase kit (Bio-Rad, CA, USA) according to the manufacturer's instructions.

### *M. truncatula* primer collection of transcription factors

A *M. truncatula* primer collection representing all annotated *M. truncatula* transcription factors was generated using available genomic databases (Udvardi et al. 2007). This resource was constituted by 712 primer pairs corresponding to 712 putative transcription factors in *M. truncatula*, and including representatives of all the major TF families. Oligonucleotide primers were designed according to common real-time RT-PCR criteria to ensure maximum specificity and efficiency during PCR amplification of cDNA (Czechowski et al. 2004). The set of 712 putative TFs with corresponding BAC and TC accession numbers as well as sequences of the primer pairs used for amplification is listed in Table S1 of the Supplementary Material.



## Real-time PCR conditions and analysis

Polymerase Chain Reactions were performed in an optical 384-well plate with an ABI PRISM<sup>®</sup> 7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA) using SYBR<sup>®</sup> Green to monitor dsDNA synthesis. Reactions, containing 2.5 µl 2X SYBR<sup>®</sup> Green master mix reagent (Applied Biosystems, Foster City, CA, USA), 0.5 µl cDNA and 200 nM of each gene specific primer in a final volume of 5 µl, were incubated at: 50°C for 2 min; 95°C for 10 min; and 45 cycles of 95°C for 15 s and 60°C for 1 min. Immediately after thermal cycling, a dissociation curve was performed by heating PCR products from 60°C to 95°C, to reveal the presence of non-specific amplicons including primer dimerization.

Data were analysed using the SDS 2.0 software (Applied Biosystems). All amplification plots were analysed with a threshold of 0.3 to obtain Ct values. PCR efficiency (E) was estimated using LinReg software with data obtained from the exponential phase of each individual amplification plot and the equation  $(1 + E) = 10^{\text{slope}}$  (Ramakers et al. 2003). To profile TFs in the three seed tissues, we used iCycler<sup>™</sup> IQ thermal cycler with 5 µl of 2X IQ SYBR Green Supermix detection (Bio-Rad, Hercules, USA), 100 nM of each gene specific primer, 2 µl of cDNA in a final volume of 10 µl. Data were analysed using iCycler<sup>™</sup> IQ Optical system software v.3.00 (Bio-Rad, Hercules, USA).

## Selection of reference genes constitutively expressed during *Medicago truncatula* seed development realized for Real-Time RT-PCR normalization

Out of the 10 housekeeping genes [*UBIQUITIN*, *ACTIN*, *GAPDH* (*Glyceraldehyde 3 Phosphate Dehydrogenase*), *PTB* (*Polypyrimidine Tract-Binding*), *PDF2* (*Protein Phosphatase 2A Subunit A3*), *PPRre* (*pentatricopeptide repeat-containing*), *EF1 $\alpha$*  (*Elongation factor 1 alpha*), *TUBULIN*, *bHLH* (*Basic helix-loop-helix transcription factor*) and *HELICASE*] present on each plate, a selection was performed based on two methods.

Firstly, by comparison of coefficient of variation (CV) values during seed development, and secondly by analysis of gene expression stability (M) using geNORM software v3.4 (Vandesompele et al. 2002). This software computes the expression ratio between all pairwise control genes for all stages. For an ideal control gene this ratio is the same for all stages. The four genes with the lowest M values, *GAPDH* (MtC00030\_GC), *ACTIN* (TC107326), *PDF2* (TC107161; Slabas et al. 1994) and *PTB* (TC111751; Salanoubat et al. 2000) were selected as controls (Supplementary Table S2).

For qualitative analysis in seed and non seed tissues, a single housekeeping gene *MSC27*, a translationally

controlled tumor protein homolog (TC85211) (Bouton et al. 2005), was used to normalize gene expression, the objective being to estimate distribution of expression of the gene.

## Statistical analyses and clustering of TFs

To analyse gene profiles between the two biological replicates, a one-way ANOVA ( $P < 0.05$ ), using Statistica software (Statsoft, Tulsa, OK), was performed on each TF. In order to calculate relative TF expression levels, the efficiency values of each amplification reaction were taken into account using LinReg software (Ramakers et al. 2003). Amplification reactions of efficiencies lower than 1.6 were considered as missing data (approximately 11.7% of the total reactions (2332/19852)). Differences in transcript abundance during seed developmental stages were also evaluated by a one-way ANOVA test ( $P < 0.05$ ) and a Student–Newman–Keuls test for each TFs using SAS software package (SAS Institute 1999). To compare expression profiles, expression values were adjusted by a normal distribution. K-means clustering were performed using Genesis software (Sturn et al. 2002) and Cluster v3.0 (Eisen et al. 1998) with the average linkage method and visualized using TreeView v1.0.5 software.

## Combination of the real-time RT-PCR profiling data with a Mt16kOLI1 microarray dataset

To combine data from the real-time RT-PCR transcript profiling and from microarray experiments previously performed (Gallardo et al. 2007), all normalized real-time RT-PCR data were expressed in log<sub>2</sub> ratios of relative expression of the reference sample (12 DAP). Genes present on the microarrays and real-time RT-PCR data were hierarchically clustered using Cluster v3.0 software (Eisen et al. 1998) according to the average linkage method, and visualized using Java TreeView v.1.0.5. To identify genes co-expressed with different TFs, we carried out a correlation analysis between the gene expression profiles from microarray dataset and different TF profiles.

## Results

### Strategy used to identify TFs expressed during seed development

The profiles of expression during seed development of a comprehensive collection of *Medicago truncatula* TF gene sequences were determined at seven developmental stages chosen based on earlier studies of seed development (Gallardo et al. 2003, 2007). The first stage, 10 Days After



Pollination (DAP), corresponds to late embryogenesis. The next stage, 12 DAP, is the transition between embryogenesis and the onset of seed filling. The following stages correspond to the onset of vicilin and legumin expression (14 DAP and 16 DAP, respectively) and the peaks of vicilin and legumin expression (20 DAP and 24 DAP). The last stage, 36 DAP, marks the end of seed filling and the entry into quiescence associated with desiccation (Fig. 1a).

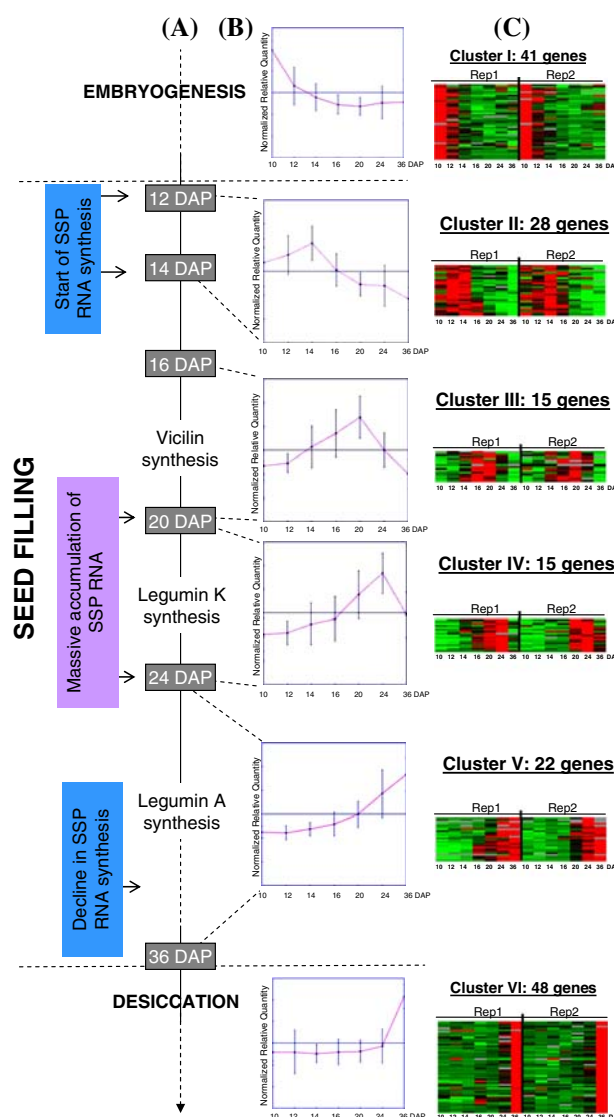
The mRNA abundances of putative TFs were determined at each stage using real-time quantitative reverse transcription PCR (real-time RT-PCR). A primer pair collection was designed to amplify each of the 712 TFs annotated in the genomic IMGAG database (the International Medicago Genome Annotation Group) and assigned to different families (Fig. 2). The representation of different TF families in the collection is similar to that observed for the *Arabidopsis* genome (Riechman 2002; Udvardi et al. 2007).

To measure expression levels accurately, we have normalized data to the geometric mean of the relative quantity of transcript for several housekeeping genes which display the least varying expression profiles throughout the stages studied (Supplementary Table S2).

The threshold cycle values of the housekeeping genes, loaded on each plate, were used to calculate the coefficients of variation (CVs) between plates. CVs calculated as the ratio between standard deviation and mean threshold cycle value, were always below 0.01 (Data not shown). More than 95% of TF transcripts (681/712) showed reproducible profiles throughout seed development between the two biological replicates (Pearson's correlation coefficient,  $r = 0.93$ ). This linear correlation was observed for expression values above  $10^{-4}$  (Supplementary Figure S1). The 72% (512/712) of genes, whose expression was detected above this threshold for at least one stage during seed development, and which were consistent between the two biological replicates, were selected for further analysis (Supplementary Table S3).

#### Analysis of TF expression profiles during seed development

To group the TFs expressed at specific stages of development, and hence to relate them to possible functions, we performed a hierarchical clustering of the profiles, along with real-time RT-PCR expression data for genes encoding the major *M. truncatula* seed storage proteins of the globulin-type. Using one-way ANOVA and Student-Newman-Keuls tests, we divided TFs into two different groups: those differentially expressed and those constitutively expressed during seed development. Approximately 67% of the TFs (343/512) are expressed throughout seed development. A k-means clustering was performed on the remaining 33% of TFs (169/512) that are differentially expressed during seed

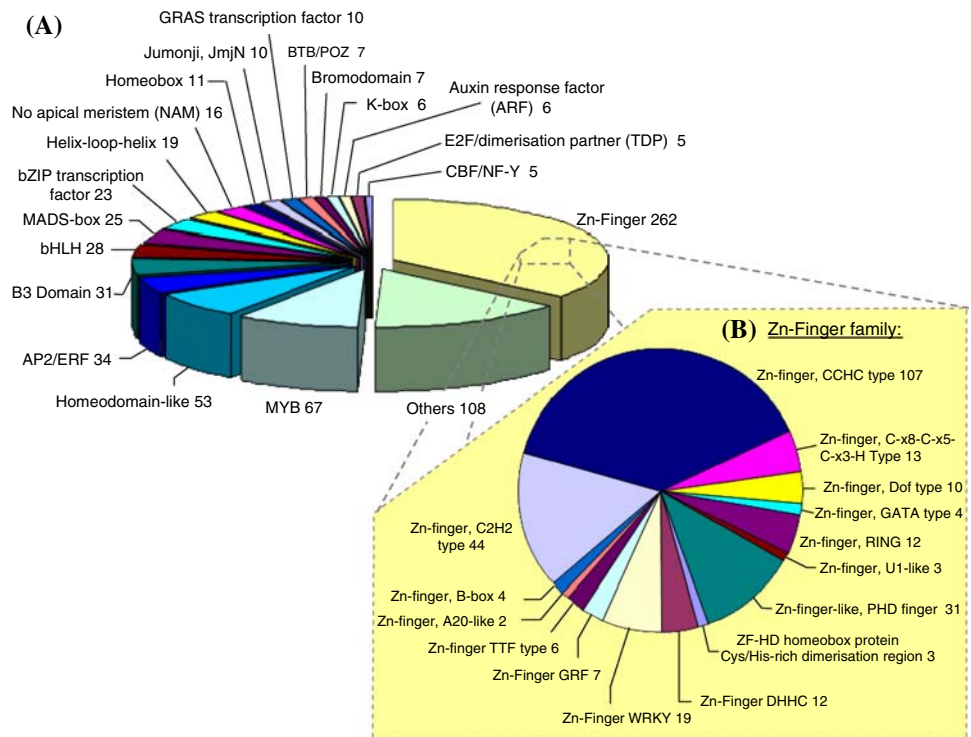


**Fig. 1** (a) Time course of *M. truncatula* seed development, with the timing of accumulation of the main Seed Storage Protein (SSP) classes. (b) Averaged Profiles of mRNA abundance during seed development for TF clusters I–VI. The  $\log_2$  of the normalized expression values (i.e. adjusted by a normal distribution) was calculated for classification into clusters. For each cluster, the median of these values at each stage is represented in graphic form ( $\pm$ SD), showing that each cluster has a characteristic peak of expression at a particular developmental stage. (c) Heat-maps for the two independent biological replicates (Rep1 and Rep2) of TF expression profiles from each cluster during seed development (10, 12, 14, 16, 20, 24 and 36 days after pollination, respectively). Down-regulation is indicated in green, up-regulation in red, normalized expression values close to the mean are in black and no detectable expression is in grey. The number of TFs present in each cluster is indicated. The identity of the TFs present in different clusters and their expression profiles are given in Supplementary Table S4

development. This cluster analysis revealed six different clusters of genes which correspond to expression profiles at different developmental stages (Fig. 1b and Supplementary Table S4). We observe major differences in TF family



**Fig. 2** Overview of the *M. truncatula* TF families present in the primer set collection. **(a)** All 712 genes were classified into TF families. The Number of TF members of each family is indicated. **(b)** The subgroups belonging to the Zinc-Finger superfamily are represented with the number of genes in each subfamily



distributions between different clusters and therefore in the proportions of different TF families expressed at different stages of seed development (Fig. 3).

The first cluster (Cluster I), corresponding to late embryogenesis, contains 41 genes which are preferentially expressed at 10 DAP. 14 of these genes are Zn-Finger family members, including sequences related to five WRKY (Eulgem et al. 2000), and notably three C2H2 type factors (Takatsuji 1999). In addition, five AP2/ERFs, four MADS and four MYB factors, one E2F-type cell proliferation factor (Sozzani et al. 2006) and a putative orthologue of the homeodomain MERISTEM LAYER 1 protein (Lu et al. 1996) were expressed at this stage. The following clusters (Clusters II–V) contain 80 TFs, with peaks of expression during seed filling (between 12 and 36 DAP). Cluster II (12–14 DAP) contains genes which are upregulated just before storage protein synthesis, marking the switch between embryogenesis and seed filling. This cluster includes candidate orthologues of genes already implicated in this transition phase such as LEAFY COTYLEDON1-LIKE (Kwong et al. 2003) and CUP-SHAPED COTYLEDON LIKE (CUC-LIKE) (Aida et al. 1999). After 14 DAP, embryo cell divisions cease, and cellular expansion phase begins, during which the principal seed reserves accumulate. The three following clusters, III, IV, and V, correspond to genes concomitantly expressed with vicilin (TC100302) (16–20 DAP), legumin K (TC100253) (20–24 DAP) and legumin A (TC100252) (24–36 DAP). In *M. truncatula*, a clear phasing in the accumulation of

transcripts for different globulin classes is observed (Gallardo et al. 2003, 2007). Vicilin mRNA synthesis starts around 14 DAP, with maximum accumulation around 20 DAP, whereas legumin K transcript accumulation begins and peaks around 4 days later than vicilin (Fig. 4). Corresponding differences in the timing of transcript accumulation of the different legumin classes, legumin K and legumin A, are also observed. TFs from clusters III to V include possible orthologues of regulators of storage protein synthesis in other species including B3 domain, bZIP, DOF and MYB family members (Fig. 4 and Supplementary Table S4).

Finally, cluster VI contains genes expressed during the desiccation phase (36 DAP). A total of 48 genes are expressed at this stage constituting the largest group of differentially-expressed TFs. Most of the TFs in this group belong to the Zn-Finger and MYB families. The bulk of the mRNA accumulated at this stage is stored and used for protein synthesis during subsequent germination, and certain of these TFs may represent potential regulators of germination performance, which is largely programmed during seed maturation (Rajjou et al. 2003).

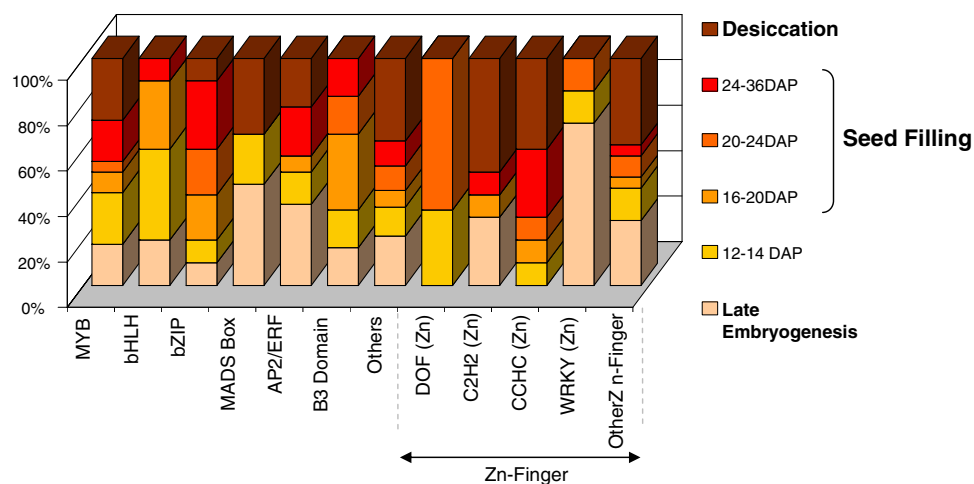
Distribution of the TFs in the three component seed tissues during seed maturation

To get a better understanding of the roles played by individual TFs, we have analyzed their expression profiles by real-time RT-PCR in seed coat, endosperm and embryo





**Fig. 3** Frequency analyses of different families of transcription factor classes expressed at different stages of seed development. The 169 TFs present in different clusters (Cluster I–VI) were classified into TF families



isolated from seed collected at 14 DAP. This stage represents the onset of seed filling and is also characterized by remobilization of endogenous reserves from endosperm to the embryo. We have also analyzed expression in roots, flowers and leaves, to assess overall cell-type specificities. The data were normalized to the gene *MSC27* (TC85211) (Bouton et al. 2005), a translationally controlled tumor protein homolog, previously validated in a *M. truncatula* comparative study using three other housekeeping genes. Among the 80 TFs preferentially expressed during seed filling, the 41 highest expressed candidates were selected for further analysis of their seed tissue specificity. We have taken as a working definition of tissue-type specificity a greater than 20-fold difference in transcript abundance between the expressing and the non-expressing cell type examined (Czechowski et al. 2004).

Three of the 41 TFs profiled by real-time RT-PCR gave no detectable transcript in any seed tissue examined, probably due to their lack of expression at 14 DAP. Six TFs were concomitantly expressed in all three seed tissues and are expressed throughout seed filling (Table 1). The remaining factors were classified according to their preferential sites of expression in seeds, (Table 1), although some are also highly expressed in other tissues.

TF genes preferentially expressed in seed coat or seed coat plus endosperm

Six TFs were expressed in the seed coat (group 1 in Table 1). They belong to the MYB, B3 domain, AP2/ERF, bZIP and MADS box families. Two of them displayed a seed coat-specific expression: a MYB R2R3 factor (1212.m00019) and an AP2/ERF member (1162.m00011). The remaining four TFs are expressed in the seed coat but also in other plant organs. They include a MADS Box factor (1274.m00002) closely related to *AGAMOUS LIKE2* (*AGL2*, Flanagan and Ma 1994) (Tblastx, score =  $1e-62$ ),

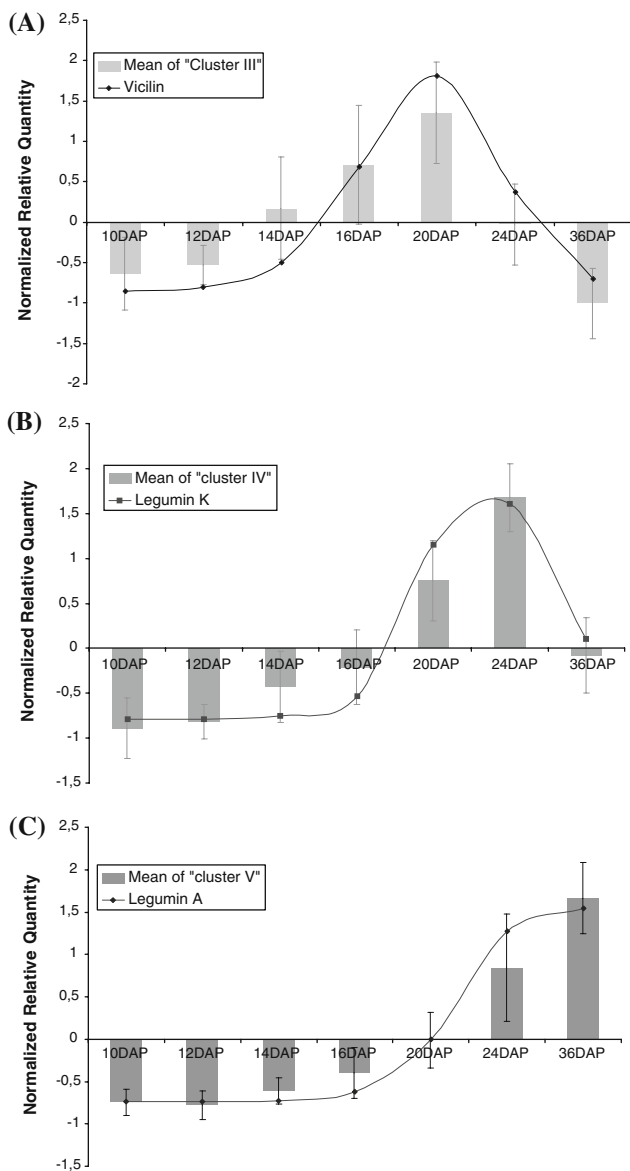
which displays a similar expression level in flower and seed coat, and a second MYB factor (727.m00020) with a R2R3 domain. Among this group were also an Auxin Response Factor (1372.m00012) closely related to auxin-responsive protein in Arabidopsis and a bZIP factor (732.m00026) which is a putative orthologue of *ATB2* (Tblastx, score =  $5e-43$ ) (Weber et al. 1995). This group is made up of genes expressed at stages which extend throughout seed filling. Our study also revealed six TFs expressed in both seed coat and endosperm of which three are seed-specific (group 2 in Table 1). This group includes essentially Zn-finger TFs (2 CCHC, 1 DOF and 1 RING type), none of which are obvious orthologues to Arabidopsis genes. The majority of these TFs are expressed early during seed filling when the seed coat and endosperm are active in supplying nutrients to the developing embryo.

TF genes preferentially expressed in embryo and/or endosperm

Six TF genes were expressed in endosperm tissue of the seed, of which three are endosperm-specific (group 3, Table 1). Whereas nine TFs were co-expressed in endosperm and embryo (group 4, Table 1), including sequences closely related to TFs of known function in Arabidopsis such as a CCAAT-box binding factor (1347.m00026) related to *LEAFY COTYLEDON 1 LIKE* (*LEC1-LIKE* or *LIL*) (Tblastx, score =  $6e-53$ ) and an AP2 factor (1376.m00008) related to *ABSCISIC ACID INSENSITIVE-4* (*ABI4*) (Tblastx, score =  $3e-33$ ) (Finkelstein et al. 1998). Only two of the nine TFs displayed seed-specific expression (*LEC1-LIKE* and *ABI4-LIKE*). Most of the TFs in this group are preferentially expressed at 12–14 DAP (i.e. at the onset of filling).

Finally, five TFs are preferentially expressed in the embryo (group 5 in Table 1). Four of these are seed- and embryo-specific, and include bHLH, bZIP and two B3





**Fig. 4** Expression profiles of the three major *M. truncatula* seed storage protein genes and of the corresponding TF cluster. The continuous line represents storage protein mRNA profiles from 10 to 36 DAP, determined by real-time RT-PCR [(a) vicilin (TC100302), (b) legumin K (TC100253) and (c) legumin A (TC100252)]. The bars represent the median ( $\pm$ SD) of normalized expression values (i.e. adjusted by a normal distribution) of TFs from the corresponding clusters (cluster III, IV and V, respectively related to vicilin, legumin K and legumin A synthesis) (Standard deviation values for the storage protein genes are close to zero). TF expression profiles related to corresponding globulin expression are available in Supplementary Table S4

domain types. Interestingly, all five embryo-expressed mRNAs of this group were expressed either in the cluster associated with vicilin (14–16 DAP) or with that associated with legumin K (20–24 DAP) genes. This group also contains a WRKY type TF sequence (920.m00011) also highly expressed in roots and flowers.

#### Identification of genes co-expressed with TFs in the seed tissues

To combine data from the real-time RT-PCR transcript profiling and from microarray experiments previously performed (Gallardo et al. 2007), we have first confirmed by 2D SDS-PAGE that the two sets of seed samples used displayed identical timing of protein accumulation (Data not shown). Further, all normalized real-time RT-PCR data were expressed as  $\log_2$  ratios of relative expression of the 12 DAP reference sample. A comparison of TF expression profiles obtained by real-time RT-PCR and microarray was then performed. Of the 169 TFs identified by real-time RT-PCR as being differentially expressed throughout seed development, approximately 40 were also present in the microarray dataset. Although the microarray dataset is less exhaustive, broadly similar expression patterns were observed for both experiments, thus validating our real-time RT-PCR results based on a different technique and normalization procedure. To propose hypotheses about the roles of individual TFs, we combined the expression profiles of the transcription factors preferentially expressed in one seed tissue, obtained by real-time RT-PCR, with profiles of seed expression of 16,000 mRNA sequences obtained by microarray-based transcriptomics (Gallardo et al. 2007). We selected 17 genes from the RT-PCR experiment which exhibit expression in only one seed tissue, disregarding any expression in other plant tissues. Of these, six genes were expressed in seed coat, six in endosperm and five in embryo. Of the microarray-derived gene list, 2,496 showed seed tissue-specific expression with less than 20% of missing data throughout seed development, 1,694 being specifically expressed in the embryo, 431 in the endosperm and 371 in the seed coat. A hierarchical cluster analysis was performed between expression profiles of these genes and those of the TFs.

To identify genes co-expressed with different TFs, gene expression profiles were considered as co-expressed when showing a correlation higher than 0.9. This analysis revealed that whereas for certain TFs many sequences are co-expressed [e.g. a WRKY factor, (920.m00011), which is co-expressed with 50 genes], for others only a few co-expressed genes were identified [e.g. a B3 domain (1299.m00024) which co-clustered with 6 genes] (Fig. 5 and Supplementary Table S5). Four TFs [MYB (1212.m00019), bZIP (732.m00026), DOF (1147.m00025) and a homeodomain factor (951.m00006)] co-clustered with only one gene. Moreover, three TFs [Zn Finger HD (783.m00019), MYB (727.m00020) and B3 domain (1372.m00012)] did not correlate with any gene expression profiles from the non-exhaustive microarray dataset.

Figure 5 shows putative functions of the co-expressed genes based on the Mapman ontology (Usadel et al. 2005). First, three of the six TFs preferentially expressed in seed



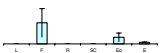
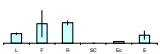
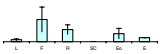
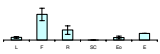
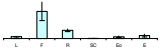
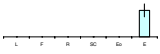
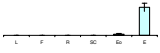
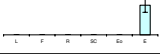
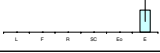
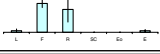
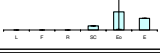
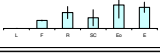
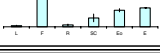
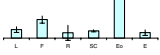
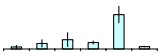
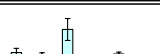

**Table 1** Expression patterns of genes identified during seed filling in the three seed tissues (Seed Coat, Endosperm, Embryo) and in non-seed plant tissues (Leaf, Flower, Root). 41 TFs are ranked according to their expression profiles into different seed tissues. TF descriptions are indicated according to tentative annotations using Blastx and the

NCBI database. Tblastx analyses using AtGI database were performed and according to their expression patterns from ATH1 22K array data, putative Arabidopsis orthologues are indicated. Corresponding clusters of TFs (according to Fig.-B1) are also indicated with their expression profiles throughout seed development

Group	ID	Description	Annotation	Cluster	Seed tissue pattern (Leaf L, Flower F, Root R, Seed Coat SC, Endosperm Eo, Embryo E)	Tblastx using AtGI DataBase	Putative <i>A. thaliana</i> Orthologues	
1- SEED COAT	1212.m00019	MYB R2R3	TC97994	16-20 DAP	SC		TC288716 GBIAAS10022.1141619094IAY19552 MYB transcription factor [Arabidopsis thaliana] (exp=-1; wgp=0; cg=0), complete Length = 1497 Score = 315 bits (688), Expect(4) = 1e-97 Identities = 121/148 (81%), Positives = 136/148 (91%)	At1g09540
	1162.m00011	AP2/ERF	TC104338	24-36 DAP	SC		TC283385 UPIBBM_ARATH (Q6PQQ4) Protein BABY BOOM (AIBBM), complete Length = 1756 Score = 454 bits (995), Expect(4) = e-130 Identities = 188/233 (80%), Positives = 206/233 (88%)	At5g17430
	1274.m00002	Transcription factor MADS Box,K-box	AC137823	12-14 DAP	F SC		TC293355 UPISEP2_ARATH (P29384) Developmental protein SEPALLATA2 (Agamous-like MADS-box protein AGL4), complete Length = 1309 Score = 178 bits (385), Expect(4) = 1e-62 Identities = 75/100 (75%), Positives = 90/100 (90%)	At3g02310
	727.m00020	Homeodomain-like;MYB R2R3	TC108487	12-14 DAP	F SC R		TC293983 UPIIMYB4_ARATH (Q9SZP1) Transcription repressor MYB4 (Myb-related protein 4) (AMYB4), complete Length = 1013 Score = 322 bits (698), Expect(3) = 1e-91 Identities = 122/131 (93%), Positives = 129/131 (98%)	At4g38620
	1372.m00012	Transcription factor B3;Auxin response factor Aux/IAA_ARF_dimerisation	AL381971	20-24 DAP	SC R L F		TC280496 UPIARFS_ARATH (Q8RYC8) Auxin response factor 19 (Auxin-responsive protein IAA22), complete Length = 3607 Score = 897 bits (1960), Expect(4) = 0.0 Identities = 359/472 (76%), Positives = 396/472 (83%)	At1g19220
	732.m00026	Basic-leucinezipper(bZIP)transcriptionfactor	TC94673	16-20 DAP	SC F R		TC287763 UPIQ65683_ARATH (Q65683) bZIP transcription factor ATB2, complete Length = 1314 Score = 155 bits (333), Expect(2) = 5e-43 Identities = 67/93 (72%), Positives = 85/93 (91%)	At4g34590
2- SEED COAT + ENDOSPERM	1020.m00007	Zn-finger-like, RING type	AC143340	16-20 DAP	Eo SC		TC289525 UPIATXR5_ARATH (Q8VZJ1) Histone-lysine N-methyltransferase ATXR5 (Tirhoax-related protein 5) (TRX-related protein 5) (Protein SET DOMAIN GROUP 15), complete Length = 1360 Score = 107 bits (229), Expect = 5e-2	
	1296.m00013	Basic helix-loop-helix dimerisation region bHLH	AC148528	12-14 DAP	Eo SC		NP235826 GBIAO11807.4IAAG13058.1 Hypothetical protein Length = 903 Score = 85.9 bits (182), Expect(4) = 6e-30 Identities = 35/44 (79%), Positives = 39/44 (88%)	
	721.m00015	Noapicalmeristem(NAM)protein	BF645682	12-14 DAP	Eo SC		TC284111 UPIO81790_ARATH (O81790) NAM / CUC2-like protein, complete Length = 1654 Score = 297 bits (644), Expect(4) = 2e-80 Identities = 115/166 (69%), Positives = 134/166 (80%)	
	969.m00001	Zn-finger, Doftype	TC95605	20-24 DAP	Eo F R SC		TC284700 GBIAAM61191.1121536859IAY084628 H-protein promoter binding factor-2a (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1818 Score = 165 bits (361), Expect(7) = 4e-70 Identities = 61/91 (67%), Post	
	1347.m00004	Zn-finger,CCHCtype	AC151709	16-20 DAP	R Eo F SC		NP455327 GBIAO05561.3IAAM15062.1 putative retroelement integrase [Arabidopsis thaliana] Length = 3648 Score = 75.8 bits (159), Expect(4) = 3e-40 Identities = 28/55 (50%), Positives = 38/55 (69%)	
	879.m00019	Zn-finger,CCHCtype	AC135606	24-36 DAP	Eo SC R L		NP180276 GBIAL138658.1ICAB75932.1 putative protein Length = 4020 Score = 103 bits (222), Expect(2) = 7e-22 Identities = 53/140 (37%), Positives = 83/140 (59%)	
3- ENDOSPERM	1147.m00025	Zn-finger,Doftype	AC146855	12-14 DAP	Eo		TC288872 UPIQDF56_ARATH (Q9FM03) Dof zinc finger protein DOF5.6 (AIDOF5.6), complete Length = 1562 Score = 142 bits (305), Expect = 1e-33 Identities = 49/63 (77%), Positives = 56/63 (88%)	
	783.m00019	ZF-HDhomeoboxproteinCys/His-richdimerisationregion	AC125474	12-14 DAP	Eo		TC290930 RFINP_178358.1115226993INM_126310 transcription factor (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1205 Score = 103 bits (219), Expect(2) = 2e-39 Identities = 35/50 (70%), Positives = 44/50 (88%)	
	1235.m00014	MybSt1	TC107215	24-36 DAP	Eo		TC300443 UPIQ9LVSO_ARATH (Q9LVSO) Myb-related transcription activator-like (MYB transcription factor), complete Length = 1622 Score = 202 bits (443), Expect(5) = 2e-93 Identities = 90/140 (64%), Positives = 99/140 (70%)	
	853.m00019	DNA-bindingWRKY	AC135100	12-14 DAP	Eo F R		TC286886 UPIWRK23_ARATH (Q22900) Probable WRKY transcription factor 23 (WRKY DNA-binding protein 23), complete Length = 1852 Score = 152 bits (327), Expect = 2e-36 Identities = 60/92 (65%), Positives = 69/92 (75%)	
	814.m00021	Myb,DNA-binding;Homeodomain-like	AC126794	12-14 DAP	F Eo		TC290418 RFINP_850213.1130685749INM_179882 transcription factor (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1227 Score = 191 bits (415), Expect(2) = 2e-62 Identities = 78/133 (58%), Positives = 96/133 (72%)	
	951.m00006	Helix-turn-helixmotif,lambdalikerepressor;Homeobox;Homeodomain-like	AC139525	12-14 DAP	F Eo		TC294039 UPIQ9LZRO_ARATH (Q9LZRO) Homeodomain-like protein, complete Length = 773 Score = 158 bits (339), Expect = 6e-39 Identities = 65/84 (77%), Positives = 74/84 (88%)	At5g03790
4- ENDOSPERM + EMBRYO	1347.m00026	Transcription factor CBF	AC151709	12-14 DAP	Eo E		TC296290 UPIFYB6_ARATH (Q84W66) Nuclear transcription factor Y subunit B-6 (AINF-YB-6) (Protein LEAFY COTYLEDON 1-LIKE), complete Length = 705 Score = 205 bits (443), Expect = 6e-53 Identities = 81/96 (84%), Positives = 90/96 (93%)	At5g47670
	1378.m00008	AP2/ERF	AC144729	24-36 DAP	E Eo		TC290286 UPIO81138_ARATH (O81138) AP2 domain family transcription factor homolog (AP2 domain transcription factor) (ABI4-abscisic acid-insensitive 4) (ABI4), complete Length = 1006 Score = 141 bits (302), Expect = 3e-33	At2g40220
	838.m00001	Noapicalmeristem(NAM)protein	AC131026	24-36 DAP	E F Eo		TC290097 cup-shaped cotyledon 3 [Arabidopsis thaliana] Length = 1105 Score = 218 bits (471), Expect(3) = 5e-81 Identities = 84/104 (80%), Positives = 92/104 (88%)	
	894.m00007	HSF/ETS,DNA-binding;Heatshockfactor(HSF)-type,DNA-binding	TC111052	20-24 DAP	E F Eo		NP306620 GBIAO683835.5IAAG50634.1 heat shock transcription factor, putative Length = 1047 Score = 235 bits (507), Expect(2) = 3e-81 Identities = 88/106 (83%), Positives = 98/106 (92%)	



Table 1 continued

Group	ID	Description	Annotation	Cluster	Seed tissue pattern (Leaf L, Flower F, Root R, Seed Coat SC, Endosperm Eo, Embryo E)	Tblastx using AtGI DataBase	Putative <i>A. thaliana</i> Orthologues	
4- ENDOSPERM + EMBRYO	904.m00009	Homeobox	AC137078	12-14 DAP	F Eo E		TC289709 GBIAAP37133.1137909316IAY251394 WOX1 protein (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1125 Score = 153 bits (329), Expect(3) = 5e-41 Identities = 60/71 (84%), Positives = 66/71 (92%)	
	1099.m00005	probableDNA-bindingproteinT6L1.1-Arabidopsisthaliana	AC146570	20-24 DAP	R F L E Eo		TC288659 RFINP_564944.11184091321NM_105555 DNA binding (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1473 Score = 133 bits (285), Expect(3) = 1e-59 Identities = 60/92 (65%), Positives = 70/92 (76%)	
	849.m00014	Zn-finger, RING	TC107912	12-14 DAP	F R Eo E L		TC304386 RFINP_850517.11306795331NM_180186 DNA binding (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1616 Score = 178 bits (383), Expect(6) = 3e-81 Identities = 63/83 (75%), Positives = 73/83 (87%)	
	749.m00018	AP2/ERF	TC104693	12-14 DAP	F R Eo E L		TC296179 UPIQ8LC30_ARATH (Q8LC30) AP2 domain containing protein RAP2.1 (At1g46768), complete Length = 790 Score = 138 bits (297), Expect(3) = 4e-37 Identities = 54/69 (78%), Positives = 59/69 (85%)	
	1407.m00028	AP2/ERF	TC102662	12-14 DAP	F R Eo E L		TC293150 GBIAAF16760.116560760AC010155 F3M18.21 (Arabidopsis thaliana) (exp=0; wgp=1; cg=0), complete Length = 1165 Score = 148 bits (318), Expect(2) = 7e-42 Identities = 58/67 (86%), Positives = 62/67 (92%)	
5- EMBRYO	762.m00015	Basic-helix-loop-helixdimerisationregionbHLH	AC124214	16-20 DAP	E		TC290791 RFINP_195498.31306916021NM_119946 DNA binding (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), partial (85%) Length = 952 Score = 88.6 bits (187), Expect(3) = 8e-32 Identities = 36/43 (83%), Positives = 42/43 (97%)	
	1018.m00015	TranscriptionalfactorB3	AC143338	16-20 DAP	E		NP452946 GBIAAC007266.5IAAD28893.1 hypothetical protein [Arabidopsis thaliana] Length = 2700 Score = 35.4 bits (71), Expect = 0.040 Identities = 15/62 (24%), Positives = 29/62 (46%)	
	1240.m00018	Basic-leucinezipper(bZIP)transcriptionfactor	AC149134	20-24 DAP	E		TC307477 UPIQ9LES3_ARATH (Q9LES3) Promoter-binding factor-like protein (ABA-responsive element binding protein 3) (AREB3), complete Length = 1763 Score = 181 bits (393), Expect(5) = 3e-81 Identities = 79/93 (84%), Positives = 85/93 (91%)	
	1299.m00024	TranscriptionalfactorB3	AC149131	16-20 DAP	E		TC297773 UPIQ82411_ARATH (O82411) FUSCA3, complete Length = 1066 Score = 230 bits (497), Expect(1) = 6e-67 Identities = 92/120 (76%), Positives = 107/120 (89%)	At3g26790
	920.m00011	DNA-bindingWRKY	BI309213	20-24 DAP	F R L E		TC291152 UPIWRK13_ARATH (Q9SVB7) Probable WRKY transcription factor 13 (WRKY DNA-binding protein 13), complete Length = 922 Score = 194 bits (418), Expect(2) = 3e-50 Identities = 75/96 (78%), Positives = 83/96 (86%)	At4g39410
6- SEED COAT + ENDOSPERM + EMBRYO	1194.m00007	Basic-helix-loop-helixdimerisationregionbHLH;Helix-loop-helixDNA-binding	BI310616	16-20 DAP	Eo E SC		TC293021 RFINP_568666.11184227021NM_124039 DNA binding (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1208 Score = 133 bits (284), Expect(3) = 6e-64 Identities = 53/71 (74%), Positives = 64/71 (90%)	
	965.m00015	Zn-finger, Dof-type	TC104194	20-24 DAP	Eo E R SC F		TC285754 UPIQD0F52_ARATH (Q93ZL5) Dof zinc finger protein DOF5.2 (AtD0F5.2), complete Length = 1804 Score = 188 bits (414), Expect(7) = 5e-79 Identities = 74/104 (71%), Positives = 82/104 (78%)	At5g39660
	878.m00030	Myb,DNA-binding;Homeodomain-like;MADF	AC135565	24-36 DAP	F E Eo SC R		TC282414 GBIAAP37748.1130725452IBT008389 At5g28300 (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 2046 Score = 137 bits (296), Expect(5) = 6e-94 Identities = 53/62 (85%), Positives = 57/62 (91%)	At5g28300
	935.m00020	Zn-finger,CCHtype;HITZn-finger	TC98775	24-36 DAP	Eo F L SC ER		TC305201 homologue to UPIQ3EBD3_ARATH (Q3EBD3) Protein At3g02065, partial (22%) Length = 509 Score = 70.1 bits (147), Expect = 9e-12 Identities = 24/33 (72%), Positives = 30/33 (90%)	
	1020.m00006	Zn-finger-like,PHDfinger;FYVE/PHDzincfinger;Zn-finger,RING	AC143340	12-14 DAP	Eo R SC F E L		TC289525 UPIATXR5_ARATH (Q8VJ11) Histone-lysine N-methyltransferase ATXR5 (Trithorax-related protein 5) (TRX-related protein 5) (Protein SET DOMAIN GROUP 15), complete Length = 1360 Score = 79.4 bits (167), Expect = 1e-14 Identit	
	742.m00011	bZIPtranscriptionfactor	TC94788	24-36 DAP	R L Eo F SC E		TC284026 UPIPGA2_ARATH (P43273) Transcription factor TGA2 (HBP-1b homologue) (AHP-1b) (AtbZIP20), complete Length = 1649 Score = 368 bits (808), Expect(2) = e-133 Identities = 155/223 (69%), Positives = 187/223 (83%)	At5g06950
No Amplification	192.m00189	MYBtranscriptionfactor-related	AC140030	24-36 DAP			TC290212 RFINP_177115.11152223621NM_105625 MYB105; DNA binding (Arabidopsis thaliana) (exp=-1; wgp=0; cg=0), complete Length = 1181 Score = 270 bits (584), Expect = 6e-72 Identities = 103/114 (90%), Positives = 114/114 (100%)	
	833.m00018	Zn-finger,CCHtype	AC130807	24-36 DAP		NO AMPLIFICATION	NP453948 GBIAAC06570.6IAAD20658.1 putative retroelement polypolyprotein [Arabidopsis thaliana] Length = 4836 Score = 309 bits (670), Expect(5) = e-99 Identities = 126/210 (60%), Positives = 157/210 (74%)	
	1158.m00005	bZIPtranscriptionfactor	BI310836	24-36 DAP		NO AMPLIFICATION	TC286600 UPIQ9SJNO_ARATH (Q9SJNO) Abscisic acid insensitive 5 (ABI5) (bZIP protein), complete Length = 1713 Score = 168 bits (367), Expect(5) = 5e-85 Identities = 77/117 (65%), Positives = 87/117 (74%)	At2g36270

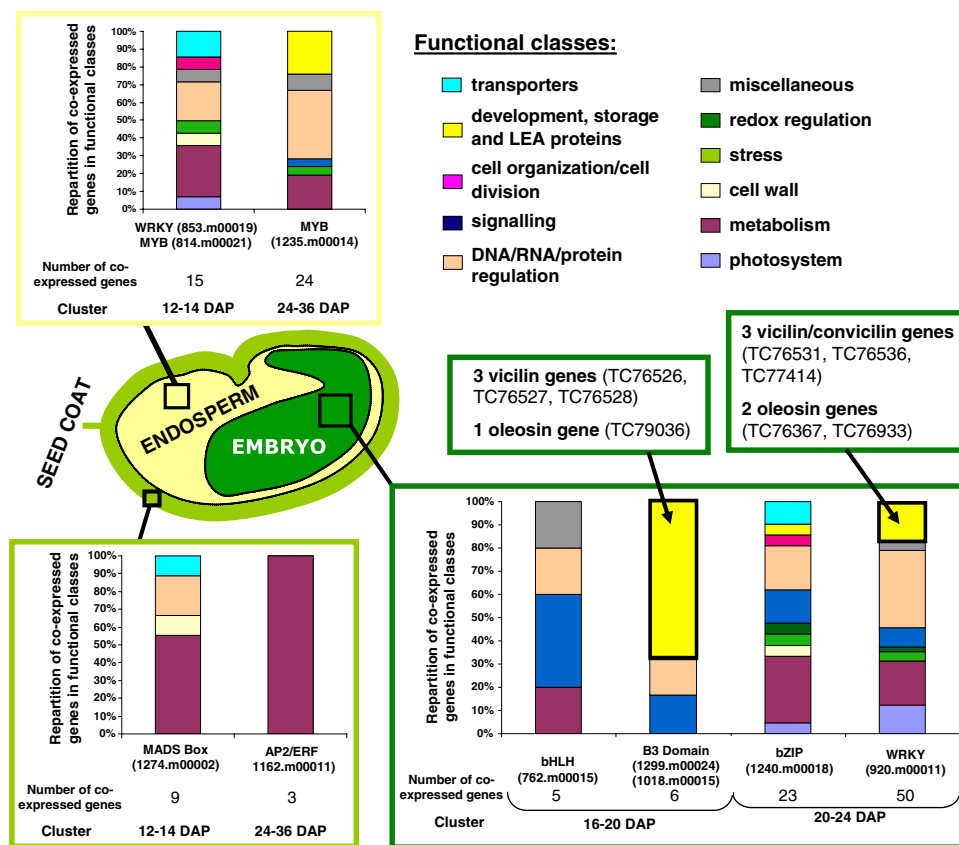
coat are co-expressed with genes involved in amino acid and secondary metabolism. Two of the six TFs expressed in endosperm (WRKY, 853.m00019 and MYB, 814.m00021) are co-expressed with the same set of genes, mainly related to sugar metabolism (e.g. a sucrose synthase, TC85400),

signalling and transport, and one (1235.m00014) is co-expressed with genes of lipid metabolism and late embryogenesis abundant proteins.

Of the five TFs preferentially expressed in the embryo, two groups could be identified. One is composed of two TFs







**Fig. 5** Functional classification of genes co-expressed with different transcription factors in the three seed tissues (embryo, endosperm and seed coat). All TFs expressed in one seed tissue at 14 DAP are represented along with the corresponding cluster of co-expressed genes (microarray dataset from Gallardo et al. 2007). The number of co-expressed genes is indicated along with the cluster, and the annotation into functional classes according to Mapman (Usadel et al. 2005). Co-expressed TFs having the same co-expressed genes are listed together. Seven TFs have not been represented in this figure: for

three (783.m00019, 727.m00020 and 1372.m00012), there were no co-expressed gene profiles in this microarray dataset and four TFs, (1212.m00019, 732.m00026, 1147.m00025 and 951.m00006), are only co-expressed with one target. A list of co-expressed genes belonging to the “development, storage and LEA proteins” functional class is detailed for three TFs (1299.m00024, 1018.m00015 and 920.m00011). The complete list of co-expressed genes (with TC and name) for each TF is given in Supplementary Table S5

(bZIP, 1240.m00018 and bHLH, 762.m00015) associated with numerous biological processes (e.g. metabolism, DNA/RNA and protein regulation, cell organization, transport) but not with seed storage proteins. The other group contains three TFs co-expressed with seed storage protein genes: a WRKY factor (920.m00011) and two B3 domain factors, (1018.m00015 and 1299.m00024).

## Discussion

Large scale profiling of transcription factors during seed development

This study represents the first high throughput real-time quantitative RT-PCR profiling of legume transcription factors throughout a developmental process and the most comprehensive profiling study of TFs in developing seeds.

In total, of 712 *M. truncatula* putative TFs examined, 512 (72%) were expressed at least at one stage during seed development. This result supports earlier observations that a majority of TFs are expressed constitutively throughout the plant but at very low levels in most cell types (Walling et al. 1986) which may be overlooked in DNA microarray analyses (Chen et al. 2002; Czechowski et al. 2004). The high sensitivity and broad linear response range of this technology has allowed us to identify the low abundance TF transcripts including those that could be associated with developmental transitions (Putterill et al. 1995; Long et al. 1996; Mayer et al. 1998). However, among the 512 TFs expressed during seed development, 343 were not differentially expressed from 10 to 36 DAP. The remaining 169 were classified into six clusters covering major developmental stages. Based on profiling data, TFs were associated with late embryogenesis processes (41 TF genes); the transition between embryogenesis and seed filling (28); the



synthesis of vicilin (15), legumin K (15) and legumin A (22), and with desiccation (48) (Figs. 1, 4).

We have carried out a more detailed analysis of 41 TFs which showed the highest relative expression, on the assumption that they are less likely to be substitutable by other genes, and that functions would be more readily attributable to them. First, by determining the seed tissue distribution of the 41 TF transcripts, 17 were found to be preferentially expressed in a particular seed tissue at one stage (14 DAP). Second, a comparison of this sub-group of 17 TFs with the expression profiles of 16,000 *M. truncatula* genes, (Mt16kOLI1 microarray dataset, Gallardo et al. 2007), allows us to propose hypotheses about biological roles for these factors.

A comparison between expression profiles of these *M. truncatula* TFs and their putative orthologues in Arabidopsis suggests conserved and distinct elements of transcriptional regulation during seed development.

#### Conservative transcriptional regulation between legume and Arabidopsis seed development

Out of 41 TFs, we found 19 putative orthologues in Arabidopsis according to amino acid sequence alignments. Using the ATH1 22K array data (www.geneinvestigator.ethz.ch), we observed that the majority of these (15/19) show similar expression patterns to the closely related *M. truncatula* sequences (Table 1 Supplementary data Table S6). Two of them are known to regulate seed size in Arabidopsis (Garcia et al. 2005) and are preferentially expressed in the seed coat (Table 1). One encodes an Auxin Response Factor-type (1372.m00012, Table 1) which could be involved in determining seed size (Schruff et al. 2006) and the second is a bZIP factor (732.m00026, Table 1), related to ATB2 (Rook et al. 1998), could influence hexose concentration, known to control seed size, by activation of cell wall invertase. Its expression during the early stages of seed filling supports a role for this factor in the control of assimilate supply to the embryo with invertases as possible targets.

Two further *M. truncatula* TF genes expressed during seed filling encode a CCAAT binding factor (1347.m00026, Table 1) closely related to *LEAFY COTYLEDON 1(LEC1)*, which coordinates embryo morphogenesis and maturation phases in Arabidopsis (Lothan et al. 1998; Kagaya et al. 2005) and an ABI4-related sequence (1376.m00008, Table 1). ABI4 is a regulator of sugar and ABA-mediated gene repression, with effects on seed maturation in combination with *LEAFY COTYLEDON* genes (Brocard-Gifford et al. 2003). Another embryo-expressed TF encodes a B3 domain transcription factor, putatively orthologous to *FUSCA3 (FUS3)* (1299.m00024, Table 1) (Tblastx, score =  $1e-67$ ), which is a direct regulator of seed storage

protein gene transcription in Arabidopsis (Parcy et al. 1997). It may be significant that this gene is co-expressed during seed filling with another B3 domain factor (1018.m00015, Table 1), as in Arabidopsis embryo two B3 domain factors, *FUSCA3* and *ABI3*, act in synergy. A hierarchical clustering of expression profiles of these B3 domain genes with the Mt16kOLI1 microarray dataset identified four encode vicilin-type seed storage proteins and an oleosin among the six co-expressed genes (Fig. 5).

#### Legume specificity in transcriptional regulation during seed development and differences in the fine regulation of legume and Arabidopsis seed storage protein gene expression

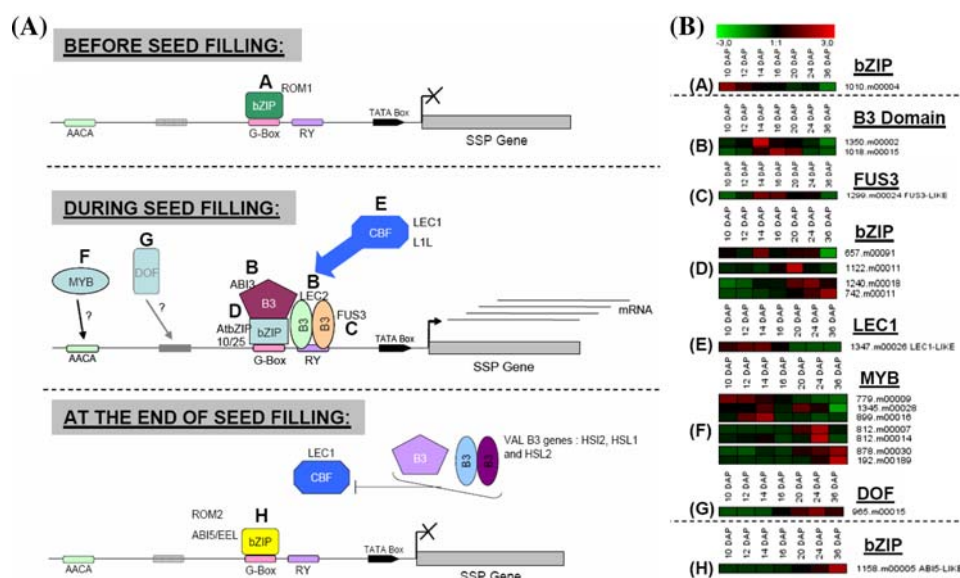
All the TFs identified as preferentially expressed in the seed coat identified have putative orthologues (i.e. sequence homology plus similar expression profile) in Arabidopsis, suggesting similar mechanisms of regulation. In contrast, for 63% (26/41) of the other *M. truncatula* TFs, there was no unambiguous corresponding orthologue in Arabidopsis (Supplementary Table S6), despite finding putative orthologues in other legumes, raising the possibility of “legume-specific” TFs (Graham et al. 2004).

As an example, three out of the five embryo TF genes lack obvious orthologues among Arabidopsis sequences or other NCBI accessions. These TFs, belonging to three different families (bHLH, B3 domain and bZIP), are expressed during seed filling with an embryo and seed specificity. Whereas in Arabidopsis the two major storage globulins and the albumins are co-ordinately transcribed (data from the ATH1 22K array available at www.geneinvestigator.ethz.ch), implying a common transcriptional regulation, in legumes vicilins and legumins accumulate successively (Meinke et al. 1981; Walling et al. 1986; Gatehouse et al. 1986; Boulter et al. 1987). Although this phasing has been widely observed in legumes, the mechanism of regulation involved has not been investigated, but may involve different TFs.

#### Models for transcriptional regulation of vicilin and legumin genes

Previous studies on other dicotyledonous species, essentially on Arabidopsis, suggest a model for coordinate transcriptional regulation of seed storage protein synthesis (Fig. 6a). *M. truncatula* factors corresponding in tissue-specificity and timing of expression to the main players in this model have been indicated. Before seed filling storage protein promoters are thought to be repressed by a mechanism involving the bZIP factor ROM1 (Chern et al. 1996b) [see *M. truncatula* candidate gene 1010.m00004 in Fig. 6(b) A]. During seed filling, storage protein gene





**Fig. 6** (a) Regulatory elements and interacting factors implied in regulating Seed Storage Protein gene promoters: A synthetic model for the transcriptional regulation of Seed Storage Protein (SSP) in dicotyledonous plants, based on data from *Arabidopsis* and *Phaseolus vulgaris* (Chern et al. 1996a, b; Reidt et al. 2000; Diaz et al. 2002; Kroj et al. 2003; Kwong et al. 2003; Lara et al. 2003; Bensmihen et al. 2005; Kagaya et al. 2005; Suzuki et al. 2007; Tsukagoshi et al. 2007). Boxes represent conserved sequence motifs: AACA motifs,

RY boxes and G boxes. Putative transcription factors and binding sites are linked by arrows. (b) Possible *M. truncatula* orthologues identified by the data presented here are indicated with letters (from A to H) corresponding to transcription factors present in model with their corresponding names and normalized expression profiles (colour code for the expression profiles is indicated on the top right and is explained in Fig. 1)

transcription is promoted by the combined action of the B3 domain proteins *LEAFY COTYLEDON2* (*LEC2*) [Fig. 6(b) B], *FUS3* [Fig. 6(b) C] and *ABSCISIC ACID INSENSITIVE-3* (*ABI3*) [Fig. 6B, Reidt et al. 2000 and Kroj et al. 2003], the latter in a complex with a *bZIP* [Fig. 6(b) D, Lara et al. 2003]. The synthesis of these factors is in turn activated by *LEC1* and *LIL* [Fig. 6(b) E, Kwong et al. 2003; Kagaya et al. 2005].

An interaction between the AACA box and a *MYB* factor [Fig. 6(b) F] is also essential for seed storage protein gene activation in cereals (Suzuki et al. 1998) and proposed for *Arabidopsis* (Vicente-Carbajosa and Carbonero 2005), and *DOF* factors [Fig. 6(b) G] are known to regulate storage protein accumulation in cereal endosperm in association with *bZIP* factors (Yamamoto et al. 2006). Finally, at the end of seed filling, further *bZIP* factors, *ROM2* (Chern et al. 1996a) or *ABI5/EEL* (Bensmihen et al. 2005) [Fig. 6(b) H] have been implicated in shutting down seed storage protein synthesis during seed desiccation. In parallel, the *VAL B3* proteins that act to repress *LEC1* or *LIL* genes (Suzuki et al. 2007) or other B3-domain genes may repress seed filling genes in conjunction with the *CHD3* chromatin-remodelling factors (Tsukagoshi et al. 2007).

All TFs identified in the present study to be closely related to putative regulators of storage protein gene expression were expressed just before or during vicilin synthesis. The later expression of legumin genes may

therefore involve other mechanisms. This could be a specific post-transcriptional activation of a TF, or chromatin remodelling, as shown for phaseolin gene regulation (Li et al. 2001). A further possibility might be the need for an additional late-expressed factor, possibly from among the 37 TFs identified in this study (e.g. from cluster IV and/or V) (Fig. 6 and Supplementary Table S4), which are co-expressed with legumin genes.

**Acknowledgements** We are grateful to all the members of laboratory, particularly Vincent Savoie for his very valuable technical support with bioinformatics analyses; Judith Burstin and Michael Bourgeois for helpful discussions; Delphine Hericher and Françoise Moussy for useful assistance with plant growth. In addition, we would also like to acknowledge Dr. Hugues Barbier (Max Planck Institute for Plant Breeding Research MPIZ, Koeln, Germany) for advice and help with RNA extraction; Helge Küster and Christian Firnhaber (Bielefeld University, Germany) who have developed the Mt16kOLI1 microarrays for fruitful collaboration. Research was supported by the FP6 EU project Grain Legumes (FOOD-CT-2004-506223).

## References

- Aida M, Ishida T, Tasaka M (1999) Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. *Development* 126:1563–1570
- Ané JM, Zhu H, Frugoli J (2008) Recent advances in *Medicago truncatula* genomics. *Int J Plant Genomics* Article ID 256597, doi:10.1155/2008/256597



- Bensmihen S, Giraudat J, Parcy F (2005) Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during *Arabidopsis thaliana* embryo maturation. *J Exp Bot* 56:597–603
- Boulter D, Evans MI, Ellis RJ, Shirsat A, Gatehouse JA, Croy RRD (1987) Differential gene expression in the development of *Pisum sativum*. *Plant Physiol Biochem* 25:283–289
- Bouton S, Viau L, Lelievre E, Limami A (2005) A gene encoding a protein with a proline-rich domain (MtPPRD1), revealed by suppressive subtractive hybridization (SSH), is specifically expressed in the *Medicago truncatula* embryo axis during germination. *J Exp Bot* 56:825–832
- Brocard-Gifford IM, Lynch TJ, Finkelstein RR (2003) Regulatory networks in seeds integrating developmental, abscisic acid, sugar, and light signaling. *Plant Physiol* 131:78–92
- Cernac A, Benning C (2004) *WRINKLED1* encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. *Plant J* 40:575–85
- Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 11:113–116
- Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitman SA (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14:559–574
- Chern MS, Bobb AJ, Bustos MM (1996a) The regulator of MAT2 (ROM2) protein binds to early maturation promoters and represses PvALF-activated transcription. *Plant Cell* 8:305–321
- Chern MS, Eiben HG, Bustos MM (1996b) The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. *Plant J* 10:135–148
- Cook RD (1999) *Medicago truncatula*—A model in the making! *Curr Opin Plant Biol* 2:301–304
- Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root and shoot specific genes. *Plant J* 38:366–379
- De Folter S, Busscher J, Colombo L, Losa A, Angenent G (2004) Transcript profiling of transcription factor genes during silique development in *Arabidopsis*. *Plant Mol Biol* 56:351–366
- Diaz I, Vicente-Carbajosa J, Abraham Z, Martínez M, Isabel-La Moneda I, Carbonero P (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant J* 29:453–464
- Druka A, Muehlbauer G, Druka I, Caldo R, Baumann U, Rostoks N, Schreiber A, Wise R, Close T, Kleinhofs A, Graner A, Schulman A, Langridge P, Sato K, Hayes P, McNicol J, Marshall D, Waugh R (2006) An atlas of gene expression from seed to seed through barley development. *Funct Integr Genomics* 6:202–211
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 99:10611–10616
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY super family of plant transcription factors. *Trends Plant Sci* 5:199–206
- Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998) The *Arabidopsis* abscisic acid response locus *ABI4* encodes an APETALA 2 domain protein. *Plant Cell* 10:1043–1054
- Flanagan CA, Ma H (1994) Spatially and temporally regulated expression of the MADS-box gene *AGL2* in wild-type and mutant *Arabidopsis* flowers. *Plant Mol Biol* 26:581–595
- Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol* 133:664–682
- Gallardo K, Firnhaber C, Zuber H, Hélicher D, Belghazi M, Henry C, Küster H, Thompson R (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: Evidence for metabolic specialization of maternal and filial tissues. *Mol Cell Proteomics* 6(12):2165–2179
- Garcia D, Fitz Gerald JN, Berger F (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* 17:52–60
- Gatehouse JA, Evans IM, Croy RRD, Boulter D (1986) Differential expression of genes during legume seed development. *Philosophical Transactions of the Royal Society of London. Biol Sci* 314:367–384
- Graham MA, Silverstein KA, Cannon SB, VandenBosch KA (2004) Computational identification and characterization of novel genes from legumes. *Plant Physiol* 135:1179–1197
- Hardtke CS, Berleth T (1998) The *Arabidopsis* gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J* 17:1405–1411
- Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T (2005) *LEAFY COTYLEDON1* controls seed storage protein genes through its regulation of *FUSCA3* and *ABSCISIC ACID INSENSITIVE3*. *Plant Cell Physiol* 46:399–406
- Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in *Arabidopsis*. *Development* 130:6065–6073
- Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003) *LEAFY COTYLEDON1-LIKE* defines a class of regulators essential for embryo development. *Plant Cell* 15:5–18
- Lara P, Oñate-Sánchez L, Abraham Z, Ferrándiz C, Díaz I, Carbonero P, Vicente-Carbajosa J (2003) Synergistic activation of seed storage protein gene expression in *Arabidopsis* by *ABI3* and two bZIPs related to *OPAQUE2*. *J Biol Chem* 278:21003–21011
- Li G, Chandrasekharan MB, Wolffe AP, Hall TC (2001) Chromatin structure and phaseolin gene regulation. *Plant Mol Biol* 46:121–129
- Long JA, moan EI, Medford JI, Barton MK (1996) A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* 404:66–69
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) *Arabidopsis* *LEAFY COTYLEDON1* is sufficient to induce embryo development in vegetative cells. *Cell* 93:1195–1205
- Lu P, Porat R, Nadeau JA, O'Neill SD (1996) Identification of a meristem L1 layer-specific gene in *Arabidopsis* that is expressed during embryonic pattern formation and defines a new class of homeobox genes. *Plant Cell* 8:2155–2168
- Mayer KF, Schoof H, Haecker A, Lenhard M, Jurgens G, Laux T (1998) Role of *WUSCHEL* in regulating stem cell fate in *Arabidopsis* shoot meristem. *Cell* 95:805–815
- Meinke DW, Chen J, Beachy RN (1981) Expression of storage-protein genes during soybean seed development. *Planta* 153:130–139
- Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J (1995) Regulation of gene expression programs during *Arabidopsis* seed development: roles of the *ABI3* locus and of endogenous abscisic acid. *Plant Cell* 6:1567–1582
- Parcy F, Valon C, Kohara A, Miséra S, Giraudat J (1997) The *ABSCISIC ACID-INSENSITIVE3*, *FUSCA3*, and *LEAFY COTYLEDON1* loci act in concert to control multiple aspects of *Arabidopsis* seed development. *Plant Cell* 9:1265–1277
- Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* 80:847–857





- Rajjou L, Gallardo K, Debeaujon I, Vandekerckhove J, Job C, Job D (2003) The effect of  $\alpha$ -amanitin on the arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiology* 134:1598–1613
- Ramakers C, Ruijter JM, Lekan Deprez RH, Moorman AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 339:62–66
- Reidt W, Wohlfarth T, Ellerström M, Czihal A, Tewes A, Ezcurra I, Rask L, Bäumlein H (2000) Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the *FUS3* gene product. *Plant J* 21:401–408
- Riechman JL (2002) Transcriptional regulation: a genomic overview (September, 30, 2002). In Somerville CR and Meyerowitz EM (eds) *Arabidopsis Book*, American Society of Plant Biologists, Rockville, MD (<http://www.aspb.org/publications/Arabidopsis/>)
- Riechman JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* 3:423–434
- Rook F, Gerrits N, Kortstee A, van Kampen M, Borrias M, Weisbeek P, Smeekens S (1998) Sucrose-specific signalling represses translation of the Arabidopsis *ATB2* bZIP transcription factor gene. *Plant J* 15:253–263
- Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* 14:1191–1206
- Salanoubat M, Lemcke K, Rieger M, Ansorge W, Unseld M, Fartmann B, Valle G, Blocker H, Perez-Alonso M, Obermaier B et al (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408:820–822
- Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. *FEBS Lett* 579:4666–4670
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ (2006) The *AUXIN RESPONSE FACTOR 2* gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133:251–61
- Slabas AR, Fordham-Skelton AP, Fletcher D, Martinez-Rivas JM, Swinhoe R, Croy RR, Evans IM (1994) Characterisation of cDNA and genomic clones encoding homologues of the 65 kDa regulatory subunit of protein phosphatase 2A in *Arabidopsis thaliana*. *Plant Mol Biol* 26:1125–1138
- Sozzani R, Maggio C, Varotto S, Canova S, Bergounioux C, Albani D, Cella R (2006) Interplay between Arabidopsis activating factors E2Fb and E2Fa in cell cycle progression and development. *Plant Physiol* 140(4):1355–1366
- Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data. *Bioinformatics* 18:207–208
- Suzuki A, Wu CY, Washida H, Takaiwa F (1998) Rice MYB protein OSMYB5 specifically binds to the AACA motif conserved among promoters of genes for storage protein glutelin. *Plant Cell Physiol* 39:555–559
- Suzuki M, Wang HH, McCarty DR (2007) Repression of the LEAFY COTYLEDON1/B3 regulatory network in plant embryo development by *VPI/ABSCISIC ACID INSENSITIVE 3-LIKE B3* genes. *Plant Physiol* 143:902–11
- Takatsuji H (1999) Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. *Plant Mol Biol* 39:1073–1078
- Tsukagoshi H, Morikami A, Nakamura K (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc Natl Acad Sci USA* 104:2543–2547
- Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, Zhang JY, Benedito V, Hofer JM, Chueng F, Town CD (2007) Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol* 144:538–549
- Unger E, Parsons RL, Schmidt RJ, Bowen B, Roth BA (2002) Dominant negative mutants of Opaque2 suppress transactivation of a 22-kD zein promoter by Opaque2 in maize endosperm cells. *Plant Cell* 5:831–841
- Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D, Scheible WR, Gibon Y, Morcuende R, Weicht D, Meyer S, Stitt M (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol* 138:1195–1204
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3
- Vicente-Carbojosa V, Carbonero P (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int J Dev Biol* 49:645–651
- Walling L, Drews GN, Goldberg RB (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci USA* 83:2123–2127
- Weber H, Borisjuk L, Heim U, Buchner P, Wobus U (1995) Seed coat-associated invertases of fava bean control both unloading and storage functions: cloning of cDNAs and cell type-specific expression. *Plant Cell* 7:1835–1846
- Yamamoto MP, Onodera Y, Touno SM, Takaiwa F (2006) Synergism between RPBFB Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol* 141:1694–1707



### **3. SELECTION DE FACTEURS DE TRANSCRIPTION IMPLIQUES DANS LE DEVELOPPEMENT DE LA GRAINE DE *M. TRUNCATULA*.**

#### **3.1 Comparaison des données d'expression obtenues en qRT-PCR avec celles obtenues par puce à ADN Affymetrix**

Une étude comparative entre les résultats d'expression des TFs obtenus en qRT-PCR et ceux obtenus avec les données "puce à ADN Affymetrix" (Benedito *et al.* 2008) a été réalisée.

Une recherche d'homologie de séquence a d'abord été faite entre les séquences des 169 TFs et celles des 15k gènes différentiellement exprimés dans la graine, identifiés respectivement en qRT-PCR et sur puce à ADN. Ensuite les profils d'expression des gènes homologues (score  $< 1e^{-100}$ ) obtenus entre ces deux méthodes ont été comparés.

Parmi les 169 TFs identifiés en qRT-PCR, 66 gènes présentant la même séquence se sont retrouvés dans les données Affymetrix, ce qui représente environ 37% des TFs identifiés en PCR quantitative. Nous pouvons constater que ce sont les TFs les plus fortement exprimés que l'on retrouve dans les données Affymetrix : en effet, 38 des 50 TFs avec le plus fort niveau d'expression de l'expérience en qRT-PCR sont retrouvés dans les données Affymetrix, soit 76%.

Ce résultat est cohérent avec la plus grande sensibilité de la qRT-PCR par rapport aux techniques d'hybridation sur puces à ADN. Grâce à cette technologie, nous avons pu identifier des TFs faiblement exprimés au cours du développement de la graine. Dans la plupart des cas, ces gènes étaient présents dans les 51k gènes de la puce mais en raison de leur faible niveau d'expression, ils n'ont pas été détectés ou de manière non reproductible et donc non retenus dans l'analyse finale. Il est important de souligner que pour les gènes présents dans les deux analyses des profils d'expression ont été observés validant ainsi le choix des gènes de contrôle utilisés pour normaliser les données de PCR quantitative.

#### **3.2. Sélection de candidats expressionnels pour l'approche de validation fonctionnelle**

##### **3.2.1 Critères de sélection**

Suite aux travaux réalisés précédemment dans ce chapitre, nous avons identifié des facteurs de transcription potentiellement impliqués dans le processus de remplissage de la graine. En vue de réaliser des analyses fonctionnelles, nous avons restreint cette liste à une dizaine de gènes candidats. Plusieurs critères de sélection ont été définis avec par ordre de pertinence :

- le profil d'expression,
- le niveau d'expression,



- la spécificité tissulaire,
- et enfin les données bibliographiques disponibles sur ce gène.

Le pattern d'expression a été l'un de nos critères essentiels dans le choix du gène candidat. Nous avons sélectionné des gènes qui présentaient une forte expression lors de l'embryogenèse tardive donc potentiellement impliqués dans le déclenchement de la transition embryogénèse-remplissage chez *M. truncatula*. Les facteurs de transcription, dont l'expression était corrélée aux profils d'expression des protéines de réserve (essentiellement vicilines et légumine K), ont également été sélectionnés car ces TFs représentent de potentiels régulateurs directs de l'accumulation des protéines dans la graine.

Le niveau d'expression du gène a également été un critère important dans le choix. Nous avons sélectionné en priorité les facteurs de transcription qui étaient fortement exprimés pendant le développement. Nous avons émis l'hypothèse que l'intensité d'expression du gène était primordiale lors de la validation fonctionnelle. En effet, dans un cas général, une mutation affectant un gène avec un niveau d'expression élevé a moins de risque d'être complétée par un autre gène endogène.

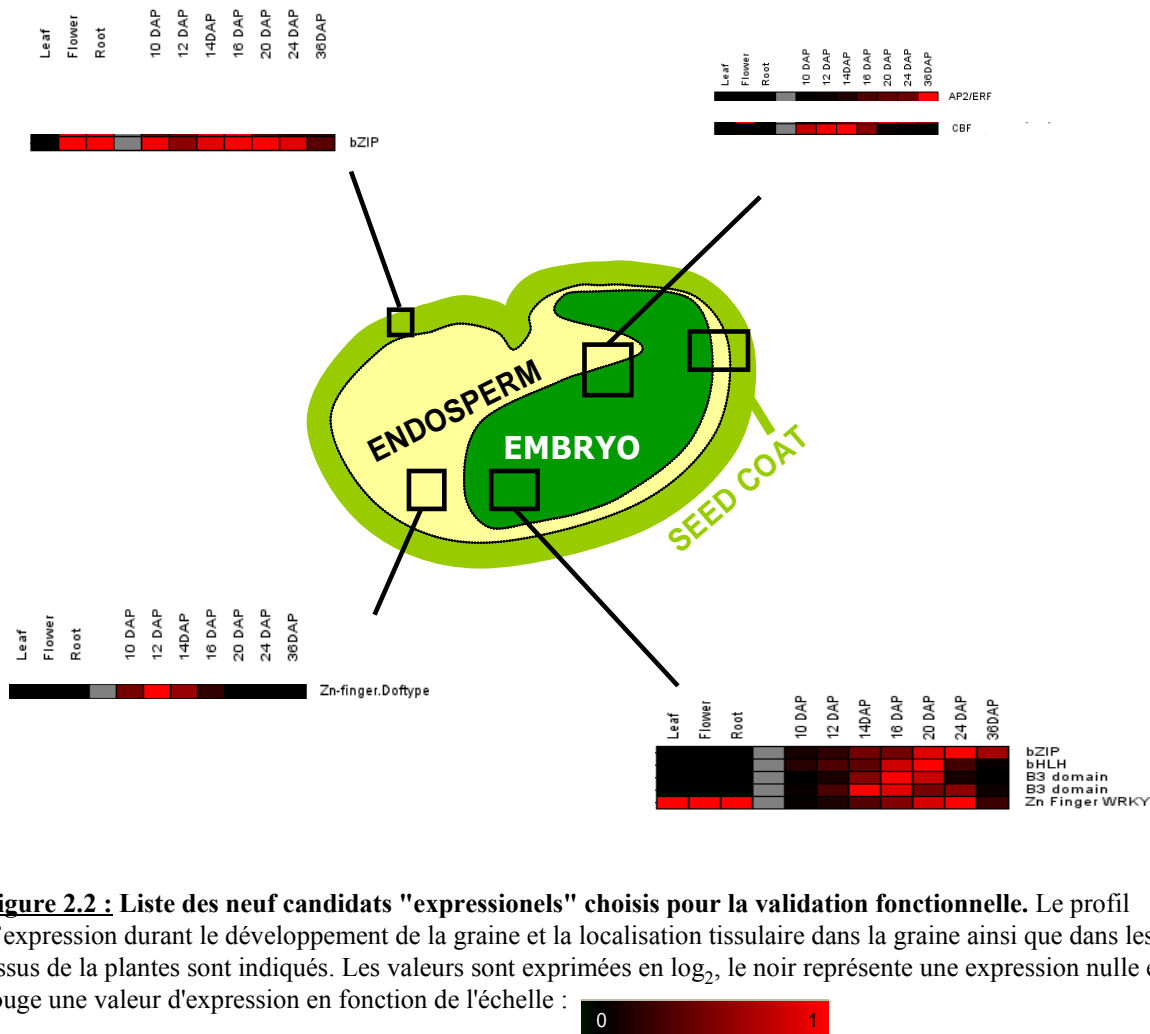
Le troisième critère a été la spécificité tissulaire du gène. Nous avons privilégié les gènes exprimés de manière spécifique dans la graine. Les candidats exprimés spécifiquement dans l'embryon, lieu de synthèse et stockage des protéines de réserve, ont tous été sélectionnés. Cependant, nous avons également retenu des facteurs de transcription exprimés dans les autres tissus de la graine (téguments et albumen) afin de mettre en évidence leur rôle ainsi que le rôle de ces tissus au niveau transcriptionnel dans le processus de développement.

Enfin, les données bibliographiques nous ont permis de sélectionner des gènes connus chez d'autres espèces pour être essentiels lors des processus de développement de la graine. Cependant, nous avons également sélectionné des gènes dont aucune donnée bibliographique n'était disponible voir même des gènes qui ne présentaient pas d'homologie de séquence avec ceux d'*A. thaliana* et donc potentiellement spécifiques aux légumineuses.

### 3.2.2 Candidats sélectionnés chez *M. truncatula*

En fonction de ces quatre critères non exclusifs de sélection, nous avons retenu neuf facteurs de transcription candidats pour une analyse fonctionnelle (Figure 2.2).

Tout d'abord, quatre sont des homologues de gènes connus chez d'autres plantes pour jouer un rôle lors le développement de la graine. Un facteur bZIP (732.m00026) qui est fortement exprimé dans les fleurs et dans les téguments des graines en développement, ce gène est proche du gène *ATB2* connu pour réguler les invertases de la paroi cellulaire (Rook *et al.* 1998). De nombreux travaux ont mis en évidence l'influence de la concentration en hexose sur les processus de division cellulaire et de différenciation dans la graine (Ohto *et al.* 2005 ou pour revue Weber *et al.* 2005). Ce gène pourrait donc intervenir dans le contrôle d'apport d'assimilats à l'embryon en



**Figure 2.2 :** Liste des neuf candidats "expressionnels" choisis pour la validation fonctionnelle. Le profil d'expression durant le développement de la graine et la localisation tissulaire dans la graine ainsi que dans les autres tissus de la plantes sont indiqués. Les valeurs sont exprimées en  $\log_2$ , le noir représente une expression nulle et le rouge une valeur d'expression en fonction de l'échelle :



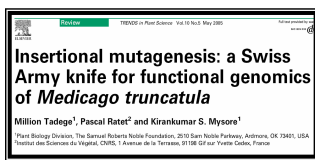
régulant les invertases et ainsi intervenir dans la régulation de la taille des graines en influençant la division cellulaire. Au niveau de l'albumen et de l'embryon, deux candidats ont été retenus : un CCAAT Binding Factor (CBF) (1347.m00026) proche du gène *LECI* qui est connu pour son rôle primordiale dans l'embryogenèse et dans la maturation de la graine d'*A. thaliana* (Lothan *et al.* 1998 ; Kwong *et al.* 2003 ; Kagaya *et al.* 2005a) ainsi qu'un facteur AP2 (1376.m00008), proche de *ABI4* qui intervient dans la régulation de l'accumulation des lipides de la graine d'*A. thaliana* (Penfield *et al.* 2006). Enfin, dans l'embryon, nous avons retenu un gène contenant un domaine B3 (1299.m00024), proche de *FUS3* qui, comme *LECI*, est un régulateur maître du développement de la graine chez *A. thaliana* (Baümlein *et al.* 1994 ; Kagaya *et al.* 2005b).

Ensuite, cinq gènes qui ne présentent pas d'homologie avec des gènes connus comme impliqués dans le développement des graines ont été sélectionnés. Un facteur Zinc Finger de type DOF (1147.m00025) exprimé spécifiquement dans l'albumen lors de la transition entre embryogenèse et remplissage. Ce gène est proche en séquence des gènes présents chez les céréales impliqués dans la régulation des protéines de réserve (Vicente-Carbajosa *et al.* 1997 ; Mena *et al.* 1998), cependant son rôle est incertain, en raison du stockage chez *M. truncatula* des protéines de réserve dans l'embryon. Il pourrait donc s'agir du gène "ancestral" à celui présent chez les monocotylédones. Pour finir, quatre gènes exprimés dans l'embryon ont également été retenus, trois sont spécifiques de ce tissu : un domaine B3 (1018.m00015) exprimé en début de remplissage, un bHLH (762.m00015) fortement exprimé à 16-20 DAP lors de la synthèse des vicilines et un bZIP (1240.m00018) fortement exprimé à 20-24 DAP lors de l'accumulation des légumineuses. Ces trois gènes pourraient être impliqués dans la régulation directe des protéines de réserve en raison de leur expression mais également de leur aptitude à fixer des éléments cis-régulateurs présents dans les gènes codants ces protéines : le domaine B3 avec le motif RY (Baümlein *et al.* 1986, Reidt *et al.* 2000, Ezcurra *et al.* 2000), le bHLH et le bZIP avec la G-box ou la B-Box (de Pater *et al.* 1993, Kawagoe *et al.* 1996)(voir revue présentée dans Introduction). Enfin le quatrième gène est un facteur Zinc Finger de type WRKY (920.m00011) exprimé pendant le remplissage de la graine au niveau de l'embryon mais également dans d'autres tissus de la plante.

### **3.3 Transfert de données chez le pois**

Afin de faciliter le transfert des connaissances sur l'espèce d'intérêt agronomique, le pois (*Pisum sativum*), une recherche d'homologie de séquence a été réalisée (tblastx) entre les 169 séquences de TFs différenciellement exprimés dans la graine de *M. truncatula* et une base de donnée privée d'ADNc de graine de pois (Hans Weber, IPK Gatersleben, Allemagne) (Künne *et al.* 2005). La recherche d'orthologue a seulement été réalisée pour les TFs qui appartenaient aux clusters 12-14 JAP (transition embryogenèse-remplissage), 16-20 JAP (synthèse des vicilines) et 20-24/24-36 JAP (synthèse des légumineuses). Vingt trois putatifs homologues ont ainsi pu être identifiés chez *P. sativum*.





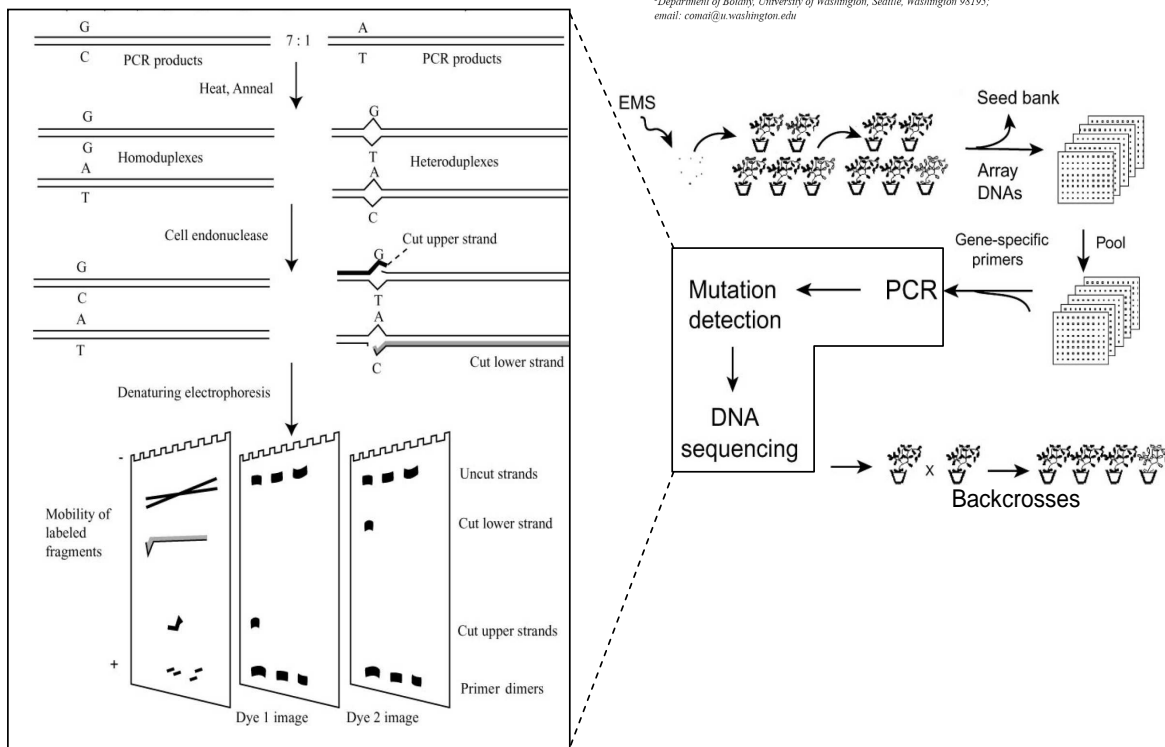
Method	Type of mutation	Average lesion frequency	Major advantages	Major disadvantages	Ease of cloning genes by forward genetics	Ease of reverse genetic screening	Flanking sequence database
Retrotransposon or <i>Tnt1</i> tagging	Gene disruption by insertion	~15 per genome	Knockout and tagging of mutated gene	Requirement for tissue culture and hence possibility of somaclonal variation; moderately difficult to generate mutant population	Easy	Easy	Yes
T-DNA tagging	Gene disruption by insertion	1-2 per genome <sup>a</sup>	Knockout and tagging of mutated gene	Requirement for transformation and tissue culture and hence possibility of somaclonal variation; difficult to generate mutant population	Easy	Easy	Yes
EMS and TILLING	Point mutations	~1000 per genome <sup>a</sup>	Allelic series for functional analysis; easy to generate mutant population	Preponderance of weak alleles	Difficult	Moderate	No
Fast neutron mutagenesis	Deletion mutations	Unknown	Knockout and deletion marker of mutated gene; easy to generate mutant population	Low frequency of allelic variation other than loss of function	Moderate	Moderate	No
RNA-induced gene silencing	Transcript knock down	Not applicable	Rapid for phenotypes accessible via hairy roots or other transient assays	Requirement for tissue culture to generate stable transformants; unpredictable penetrance	Not applicable	Easy	No

**Tableau 2.1 :** Tableau comparatif des différentes méthodes permettant de générer des mutants "perte de fonction", extrait de Tadege *et al.*, 2005.

Ann. Rev. Plant Biol. 2003. 54:375-401  
 doi: 10.1146/annurev-arplant.54.031902.135009  
 Copyright © 2003 by Annual Reviews. All rights reserved.  
 First published online as a Review in Advance on January 24, 2003

### SINGLE-NUCLEOTIDE MUTATIONS FOR PLANT FUNCTIONAL GENOMICS

Steven Henikoff<sup>1</sup> and Luca Comai<sup>2</sup>  
<sup>1</sup>Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109; email: steveh@fhcrc.org  
<sup>2</sup>Department of Botany, University of Washington, Seattle, Washington 98195; email: comai@u.washington.edu



**Figure 2.3 :** Principe du criblage TILLING (extrait de Henikoff and Comai, 2003)

Une population de plantes sont traitées à l'EMS. L'ADN est extrait et une PCR est réalisée pour amplifier le fragment du gène dans lequel on recherche une mutation. Ce produit PCR en mélange avec un fragment sauvage va être clivé par une endonucléase à l'endroit de la mutation. La position de cette mutation est identifiée par migration sur gel d'acrylamide.

## 4. ISOLEMENT DE MUTANTS POUR LES GENES CANDIDATS SELECTIONNES

### 4.1 Outils de génomique fonctionnelle disponibles chez *M. truncatula*

Plusieurs méthodes ont été développées chez les plantes pour générer des mutants "perte de fonction". On peut cependant distinguer deux grandes approches : la transgénèse et la mutagenèse. Les différentes méthodes, détaillées ci-dessous, sont comparées dans le tableau 2.1 (Li *et al.* 2002 ; Tadege *et al.* 2005).

#### 4.1.1 La mutagenèse

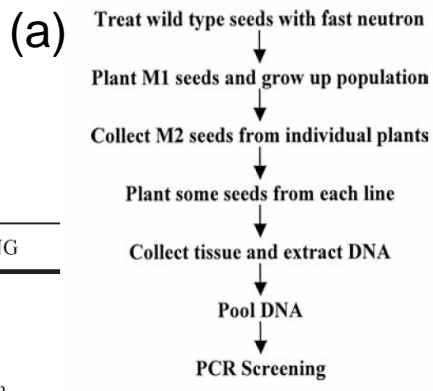
Cette méthode est adaptée à toutes les espèces de plantes et est basée sur l'action de mutagènes chimiques (de type Ethyl Méthane Sulfonate) ou physiques (rayons  $\gamma$ , rayons X, Fast Neutron, ...). Ces méthodes sont couramment utilisées pour introduire des mutations dans des génomes animaux ou végétaux.

##### 4.1.1.1 Population EMS/TILLING

L'Ethyl Méthane Sulfonate (EMS) est un mutagène chimique capable d'induire de nombreuses mutations ponctuelles de type changement d'un C/G en T/A dans un génome. Cette technique peut être utilisable pour des approches de génétique "forward" (observation d'un phénotype précis et identification du gène responsable du phénotype, approche allant du phénotype vers le gène) et de génétique "reverse" (recherche d'une mutation dans un gène choisi pour étudier son phénotype, approche partant du gène vers le phénotype). Cependant, lors de la démarche en génétique "forward", l'isolement du gène muté peut être difficile et nécessite de réaliser un clonage positionnel qui peut être long et fastidieux. Parmi les méthodes de génétique "reverse", la méthode de criblage TILLING (Targeting Induced Local Lesions IN Genomes) est utilisée (Figure 2.3) (Mc Callum *et al.* 2000). Cette technique permet l'identification d'une mutation dans une séquence nucléotidique connue. L'identification de la mutation repose sur la formation d'un hétéroduplex entre les fragments PCR de la séquence mutée et de la séquence sauvage, hétéroduplex qui sera clivé par une endonucléase spécifique des mésappariements. Ces populations existent déjà chez de nombreuses espèces végétales (arabidopsis, Greene *et al.* 2003 ; blé, Slade *et al.* 2005 ; maïs, Till *et al.* 2004 ; lotier, Perry *et al.* 2003 ; pois, Dalmais *et al.* 2008). Chez *M. truncatula*, deux populations EMS/TILLING sont actuellement disponibles : l'une à l'Université de Californie (Davis, USA) et l'autre à l'INRA UMRLEG (Dijon, FR).

##### 4.1.1.2 Population Fast-Neutron et rayons $\gamma$

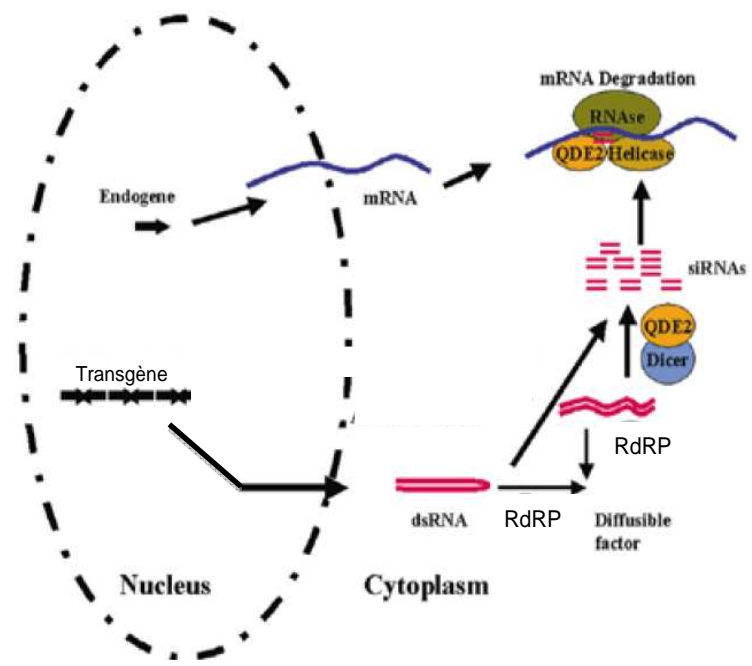
La mutagenèse physique est l'autre mutagène classique pour obtenir une population de mutants. Contrairement à l'EMS, elle ne génère pas un changement de base ponctuel mais des délétions de régions d'ADN (de quelques bases à plusieurs kb, Li *et al.* 2002) ou des réarrangements chromosomiques. C'est également un outil préférentiellement utilisé pour des approches de génétique "reverse". L'identification des plantes mutées pour un gène se fait par comparaison de taille des fragments PCR du gène d'intérêt (Figure 2.4). Chez *M. truncatula*, deux populations de Fast Neutron sont disponibles : l'une à la Noble Foundation (Ardmore, USA) et l'autre au



(b)

	Deleteagene	T-DNA/transposon	TILLING
Applicability	Wide	Limited	Wide
Developing time	Short	Long	Short
Gene preference	No	Yes	No
Tandem repeat knockout	Yes	No	No
Population size	Large	Large	Medium
Screening throughput	High	High	Medium

**Figure 2.4 :** (a) Principe du criblage d’une population de mutants « Fast Neutron » et (b) comparaison de la méthode avec le TILLING et les T-DNA/transposon (extrait de Li et Zhang, 2002).



**Figure 2.5 :** Principe de l’ARN interférence (PTGS) (extrait de Chicas *et al.* 2001).

Le double brin ARN est produit par le transgène (dsRNA), puis ce double brin va être utilisé comme matrice par une ARN polymérase ARN dépendante (RdRP). Les fragments ARN double brin sont clivés en petits ARN double brin (siRNA) par Dicer, une RNase de type III. Ces petits ARN (siRNA) vont guider un complexe dont une nucléase vers les molécules ARNm endogènes qui vont être dégradés. Cette dégradation va générer de nouveaux siRNA qui vont nourrir ce mécanisme d’inhibition de l’expression. Ce mécanisme est également accompagné de l’inhibition de l’expression du gène par méthylation de son promoteur (non représenté sur cette figure).

John Innes Center (Norwich, UK). Une population rayons Y est également disponible à l'INRA UMR-LEG (Dijon, FR) (Sagan *et al.* 2005).

#### 4.1.2 La transgénèse

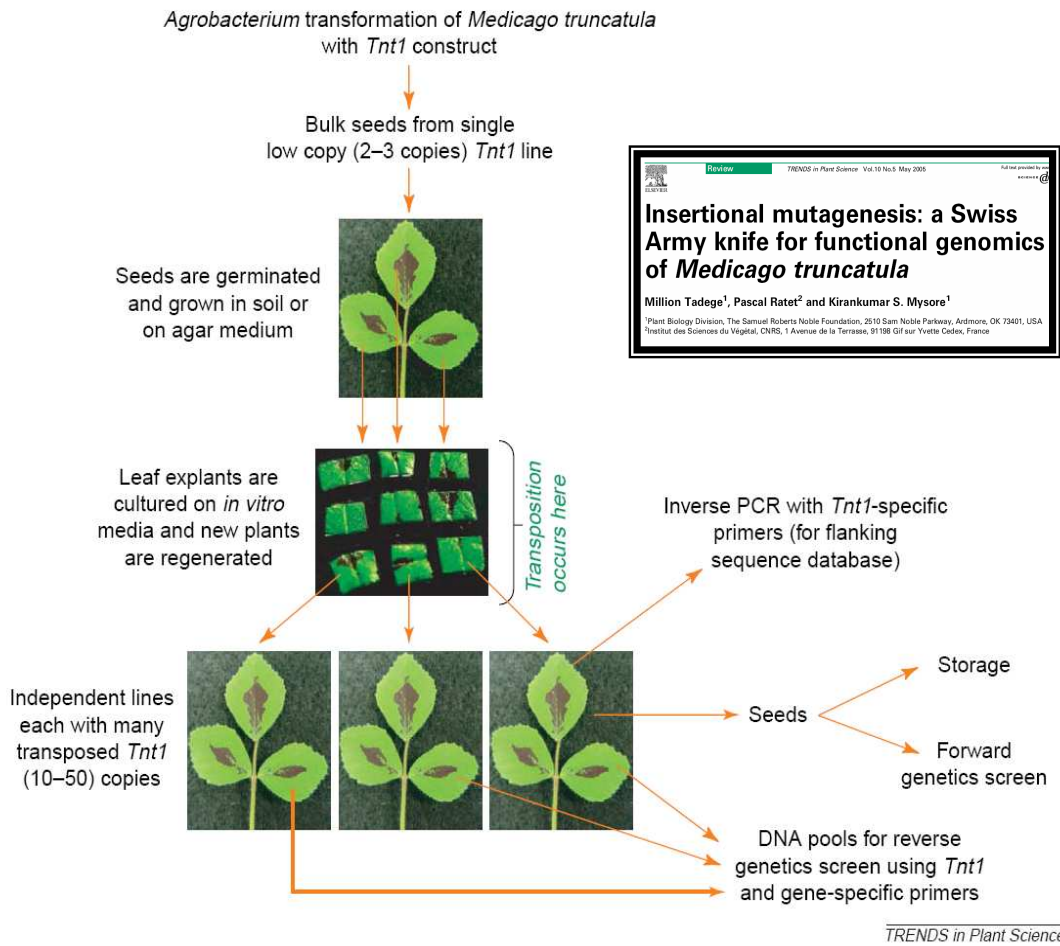
On peut distinguer différentes méthodes d'inactivation de gène par transgénèse disponibles chez *M. truncatula* : par interférence (RNAi, gene silencing ou Post-Transcriptional Gene Silencing, PTGS), par mutagenèse insertionnelle (T-DNA tagging) et par l'utilisation d'éléments transposables (transposon tagging). Ces méthodes sont couramment utilisées chez les végétaux mais nécessitent que la plante puisse être transformée et surtout régénérée *in vitro*. De plus, la régénération *in vitro* des plantes est connue pour induire de multiples mutations dans le génome des régénérants (Larkin *et al.* 1981). D'autres méthodes récentes existent également, telles que la transformation du gène d'intérêt couplé à un domaine répresseur (Hiratsu *et al.* 2003), cependant cette méthode encore peu utilisée chez les plantes ne sera pas traitée.

##### 4.1.2.1 RNA interférence (ou PTGS)

Le RNAi est un mécanisme cellulaire de régulation de l'expression des gènes. Ce mécanisme repose sur la synthèse d'un petit ARN d'une vingtaine de bases qui va se fixer sur une séquence complémentaire d'ARNm et ainsi entraîner la formation d'un double brin d'ARN. Ce complexe double brin d'ARN sera dégradé par le complexe RISC (RNA induced silencing complex) entraînant la dégradation des ARNm, le blocage de la traduction et l'arrêt de la transcription par méthylation de tous les ARNm correspondants (Figure 2.5 et pour revue Hannon *et al.* 2002). Ce mécanisme naturel peut être détourné pour inhiber l'expression d'un gène spécifique ou l'expression d'une famille entière de gènes. Il consiste à transformer la plante avec un vecteur contenant un fragment du/des gène(s) à éteindre, en construction sens et antisens. Cette méthode puissante de génétique "reverse" permet d'obtenir une variété de phénotypes allant d'une faible extinction à une extinction quasiment complète du gène. Il existe également depuis peu de temps des vecteurs permettant de réaliser ce type de transformation à haut débit conduisant à des populations de transformants (Helliwell and Waterhouse, 2003).

##### 4.1.2.2 Mutagenèse insertionnelle

Cette technique permet de muter mais également de "marquer" le génome, elle est donc employée dans des approches "forward" et "reverse". Elle consiste en l'utilisation d'une séquence d'ADN (T-DNA) qui va s'insérer aléatoirement dans le génome et donc parfois dans des gènes. L'identification des régions d'insertion peut être réalisée par PCR et/ou séquençage des régions bordantes du T-DNA. Cette technique est très largement utilisée et a fait ses preuves chez *A. thaliana*, où de nombreuses validations fonctionnelles ont été réalisées à partir de ce type de mutation. Cette ressource est disponible chez *M. truncatula* à l'Institut des Sciences du Végétal (ISV-CNRS) (Gif/Yvette, FR) (Scholte *et al.* 2002).



**Figure 2.6 :** Principe du criblage d'une population de mutants contenant le rétrotransposon *Tnt1* (extrait de Tadege *et al.*, 2005)

#### 4.1.2.3 Utilisation d'éléments transposables

Cette technique est en partie basée sur la précédente et permet de générer une grande population de mutants. Elle consiste en l'introduction dans le génome de la plante d'un transposon soit de type I : rétrotransposon qui copie son ARN et s'insère ailleurs dans le génome (transposition de type copier-coller), soit de type II : transposon à ADN qui "saute" dans le génome (transposition de type couper-coller). Ces éléments mobiles sont donc capables de s'insérer dans un nouvel endroit du génome, par exemple à chaque étape de régénération *in vitro* de la plante dans le cas spécifique du transposon Tnt1. De la même manière que précédemment, ces populations de mutants sont utilisées en génétique "forward" et "reverse" et l'identification des régions mutées se fait par PCR et/ou séquençage des bordures du transposon. Chez *M. truncatula*, des populations de ce type ont été réalisées grâce au rétrotransposon Tnt1 du tabac sur des génotypes R108 à l'ISV-CNRS (Figure 2.6) (d'Erfurth *et al.* 2003) et plus récemment sur le génotype A17 dans le cadre d'une collaboration entre l'ISV-CNRS (Gif/Yvette, FR) et la Noble Foundation (Ardmore, USA).

### **4.2 Mutants obtenus**

Afin d'identifier des plantes mutantes pour nos gènes candidats (paragraphe 4.2 et Figure 2.2), nous avons criblé les différentes populations de mutants disponibles chez *M. truncatula* précédemment décrites : les populations EMS/TILLING présentes au laboratoire (INRA UMRLEG), les populations "fast-neutron" du John Innes Center (Norwich, UK) et de la Noble Foundation (Ardmore, USA), ainsi que les populations Tnt1 de l'ISV-CNRS (Gif sur Yvette, FR) et de la Noble Foundation (Ardmore, USA).

Nous avons identifié deux à neuf mutants pour la quasi-totalité de nos gènes d'intérêt grâce à la population EMS/TILLING présente au laboratoire (INRA UMRLEG, Dijon). En revanche, aucun mutant n'a pu être identifié pour le gène *FUS3-LIKE*. Un tableau récapitulatif (tableau 2.2) présente les différentes mutations pour les différents gènes, ainsi que le type de mutation (substitution d'un acide aminé ou apparition d'un codon stop prématuré). Ce tableau renseigne également dans le cas d'une substitution d'acide aminé sur le changement ou non de propriété physico-chimique de l'acide aminé et sur la prédiction faite par l'algorithme SIFT (disponible sur le site web : <http://blocks.fhcrc.org/sift/SIFT.html>). Cet algorithme prédit l'effet de la substitution d'un acide aminé sur la fonction de la protéine en prenant en compte la conservation de séquence protéique chez les plantes mais également chez d'autres espèces où la séquence est disponible (Ng et Henikoff, 2003). Cette analyse est basée sur le fait que si un acide aminé est conservé dans toutes les séquences protéiques orthologues alors il est essentiel. Afin d'éliminer la majorité des mutations autres que dans le gène d'intérêt, deux rétrocroisements sont systématiquement réalisés avec le génotype sauvage (A17), suivis d'une à deux autofécondations. Deux rétrocroisements ont été réalisés pour sept mutants et un seul rétrocroisement a été réalisé pour 2 mutants, dans le temps imparti pour la thèse. L'état d'avancement de purification du fond génétique est détaillé dans le tableau 2.2.

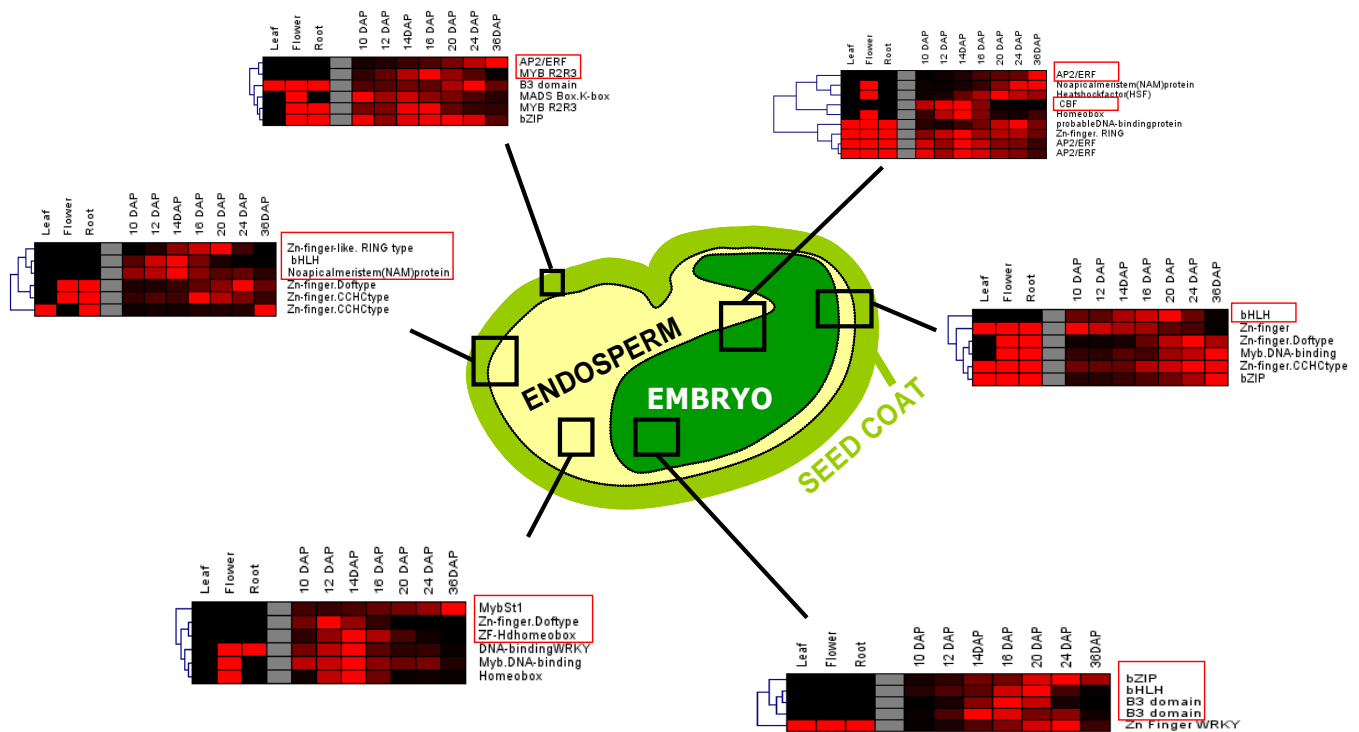
Gènes	Identifiant TIGR	Population Mutants	Type de mutation	Remarques	SIFT Prediction	Etat d'avancement du matériel
ATB2-LIKE	732.m00026	EMS (UMRLEG)	D 54 N T 78 I	Acide à neutre Hydrophile à Hydrophobe	Affect protein function	BC1
					Affect protein function	BC1
bHLH	762.m00015	EMS (UMRLEG)	C297 STOP E 57 K V 117 I D 217 N G 243 E P 121 S G 110 R	Protéine tronquée Acide à Basique même propriété Acide à Neutre Neutre à Acide Hydrophobe à Hydrophile Neutre à Basique		BC1
					Tolerated	
					Tolerated	
					Tolerated	
					Tolerated	
					Tolerated	
DOF2	1147.m00025	EMS (UMRLEG)	W 59 STOP S 222 F G 281D S 146 F R 225 Q R 264 K T 35 I G 317 E L 323 F	Protéine tronquée Hydrophile à Hydrophobe Neutre à Acide Hydrophile à Hydrophobe Basique à Neutre même propriété Hydrophile à Hydrophobe Neutre à Acide même propriété		BC2
					Tolerated	
					Tolerated	
					Tolerated	
					Tolerated	
					Tolerated	
					Affect protein function	mutation non retrouvée
					Tolerated	
					Tolerated	
WRKY	920.m00011	EMS (UMRLEG)	V 236 M S 361 F	même propriété Hydrophile à Hydrophobe	Affect protein function	BC1
					Affect protein function	BC1
bZIP	1240.m00018	EMS (UMRLEG)	G 210 A D 208 N	Neutre à Acide Acide à Neutre	Tolerated	BC1
					Tolerated	
LEC1	1347.m00026	EMS (UMRLEG)	E 44 K L 24 F S 123 L	Acide à Basique même propriété Hydrophile à Hydrophobe	Affect protein function	mutation non retrouvée
					Affect protein function	mutation non retrouvée
					Affect protein function	BC2S1
FUS3-LIKE	1299.m00024	Pas de mutants				
ABI4-LIKE	1376.m00008	Criblage en cours				
B3 Domain	1018.m00015	Criblage en cours				
AP2/ERF	1054.m00007	Tnt1 (NOBLE FOUNDATION)		insertion transposon		
AP2/ERF	1207.m00023	Tnt1 (ISV-CNRS)		insertion transposon		

**Tableau 2.2 : Liste des gènes candidats "expressionnels" ainsi que la nature et position des mutations identifiés.** Le changement de propriété physico-chimique de l'acide aminé substitué dans le cas d'une mutation ponctuelle est indiqué, ainsi que la prédiction SIFT sur l'effet de cette substitution sur la fonction de la protéine. BC1, premier rétrocroisement avec le génotypage sauvage A17; BC2, deuxième rétrocroisement avec le génotype sauvage; BC2S1, deuxième rétrocroisement suivi d'une autofécondation.

L'approche de transgénèse par interférence, plus couteuse en temps, a également été entreprise pour un seul candidat, le gène de type *LECI-LIKE* (voir détail dans Chapitre 3).

En revanche, aucune plante mutante pour nos gènes candidats n'a été identifiée dans les autres populations de mutants (rayons  $\gamma$ , Tnt1, fast-neutron, ...). La disponibilité des séquences bordantes pour les populations Tnt1 nous a permis d'identifier *in silico* des mutants pour deux autres facteurs de transcription qui n'avaient pas été retenus dans notre premier choix. Il s'agit de deux gènes à domaine AP2/ERF l'un (1207.m00023) avec un pic d'expression lors de l'embryogenèse tardive (10 DAP), l'autre (1054.m00007) exprimé en fin de développement de la graine (à 24-36 DAP) (Tableau 2.2).





**Figure 2.7 :** Expression tissulaire des facteurs de transcription ayant été analysés dans les trois tissus de la graine. Les profils d'expression au cours du développement ainsi que dans les autres tissus de la plantes sont indiqués. Les neuf TFs sélectionnés comme candidats sont encadrés. Les valeurs sont exprimés en log2, le noir représente une expression nulle et le rouge une valeur d'expression en fonction de l'échelle :



## **5. CONCLUSION DU CHAPITRE**

Au commencement de ce travail, peu de données étaient disponibles sur l'expression des facteurs de transcription chez *M. truncatula*. En utilisant la PCR quantitative à haut débit, nous avons pu suivre l'expression de 712 facteurs de transcription au cours du développement de la graine en utilisant une technique répétable et surtout l'une des plus sensibles disponibles à ce jour. En résumé, ce travail a donc permis (i) de générer un catalogue d'expression de TFs dans la graine, (ii) d'identifier de potentiels régulateurs clefs des étapes de développement de la graine et (iii) d'isoler des mutants de ces régulateurs pour leur caractérisation fonctionnelle.

### **Synthèse des résultats**

Nos données ont permis dans un premier temps de générer l'un des plus important **catalogue d'expression de facteurs de transcription lors du développement de la graine**, allant de l'embryogenèse jusqu'à la dessiccation. Nous avons ainsi pu identifier 512 TFs exprimés dans la graine en utilisant une des méthodes les plus sensibles qui existent actuellement. Parmi ces gènes, 343 sont exprimés de manière peu contrastés tout au long du développement alors qu'au contraire 169 sont différenciellement exprimés à l'un des stades correspondant à l'embryogenèse, au remplissage ou à la dessiccation. Une analyse approfondie de l'expression d'une quarantaine de TFs associés au stade de remplissage a été réalisée et a permis d'identifier des gènes spécifiques d'un des trois tissus de la graine (Figure 2.7).

Dans un deuxième temps, ce travail nous a permis d'**identifier des gènes candidats potentiellement impliqués dans des processus clefs du remplissage de la graine**. Ces candidats "expressionnels" ont été choisis car potentiellement impliqués dans la phase de transition embryogenèse-remplissage ou en fonction de leur cinétique d'expression corrélées à celle des principales protéines de réserve, les vicilines et les légumine. Ces TFs pourraient être des régulateurs spécifiques des différentes classes de protéines de réserve et ainsi expliquer la séquentialité de leur synthèse chez les légumineuses.

Enfin des mutants, pour **huit de ces gènes considérés comme impliqués dans le développement, ont été isolés** en utilisant les différentes populations de mutants disponibles chez *M. truncatula* **et seront disponibles, après épuration du fond génétique, pour une caractérisation fonctionnelle**.

### **Perspectives de ce travail**

Ce travail a avant tout servi de base à de nombreux projets et collaborations. Il a permis d'établir une collaboration forte avec le Max Planck Institute of Molecular Physiology (A. Schlereth, MP-MPI, Golm, Allemagne) qui réalise une approche semblable chez *A. thaliana* et avec la Noble Foundation (M. Udvardi, Ardmore, USA) qui travaille sur les différents aspects transcriptionnels du développement de *M. truncatula*.



En plus des données générées, ce travail a débouché sur de nombreux projets aussi bien chez l'espèce modèle *M. truncatula* que chez la plante d'intérêt agronomique, *P. sativum* :

- Une analyse en double hybride d'une dizaine de TFs candidats identifiés grâce à cette analyse est en cours de réalisation afin d'identifier les partenaires de ces TFs (G. Angenent, Wageningen, Pays-Bas).
- Chez *M. truncatula*, 120 TFs identifiés comme putativement impliqués dans la phase de remplissage ont été cartographiés. Une première recherche de QTL/PQL a mis en évidence une colocalisation entre deux QTL/PQL de teneur en protéines de la graine et deux TFs précédemment décrits (un domaine B3 et un facteur MYB) (projet ANR QualityLegSeed, K. Gallardo).
- Chez *P. sativum*, 19 TFs identifiés comme potentiels orthologues des gènes décrits chez *M. truncatula* ont été cartographiés (Bordat *et al.* in prep). Une recherche de QTL/PQL a d'ores et déjà mis en évidence une colocalisation entre un QTL majeur de teneur en protéines de réserve et un facteur DOF (Bourgeois *et al.* in prep).
- Enfin, des projets concernant la validation fonctionnelle de certains facteurs de transcription identifiés grâce à ce travail sont également envisagés.

Le résultat de ce travail a également été utilisé dans le cadre de ma thèse. Tout d'abord des analyses pour identifier d'éventuelles régulations directes de TFs sur les promoteurs de protéines de réserve ont été initiées en utilisant la transformation transitoire. Ce projet consistait à introduire par biolistique deux vecteurs l'un capable d'exprimer le TFs candidat grâce au promoteur 35S et l'autre comportant la région promotrice des gènes codant les protéines de réserve couplée avec la séquence codante de GUS. L'interaction entre notre TFs et la région promotrice des protéines de réserve aurait été visualisée par la coloration GUS. Cependant, ce projet n'a pas abouti faute de temps nécessaire à la mise au point des conditions de biolistique sur *M. truncatula*.

Enfin toujours dans le cadre de ma thèse, ces résultats ont permis de sélectionner l'un des gènes candidats identifiés, le gène LEC1-type (1347.m00026), afin de caractériser fonctionnellement son rôle dans le développement de la graine de *M. truncatula* (Chapitre 3) et de valider un potentiel effet de ce gène sur la qualité finale du grain chez *P. sativum* (Chapitre 4).



## **CHAPITRE 3 : Les gènes de type LEC1 chez *M. truncatula***



## **CHAPITRE 3 : Les gènes de type LEC1 chez la légumineuse modèle *Medicago truncatula***

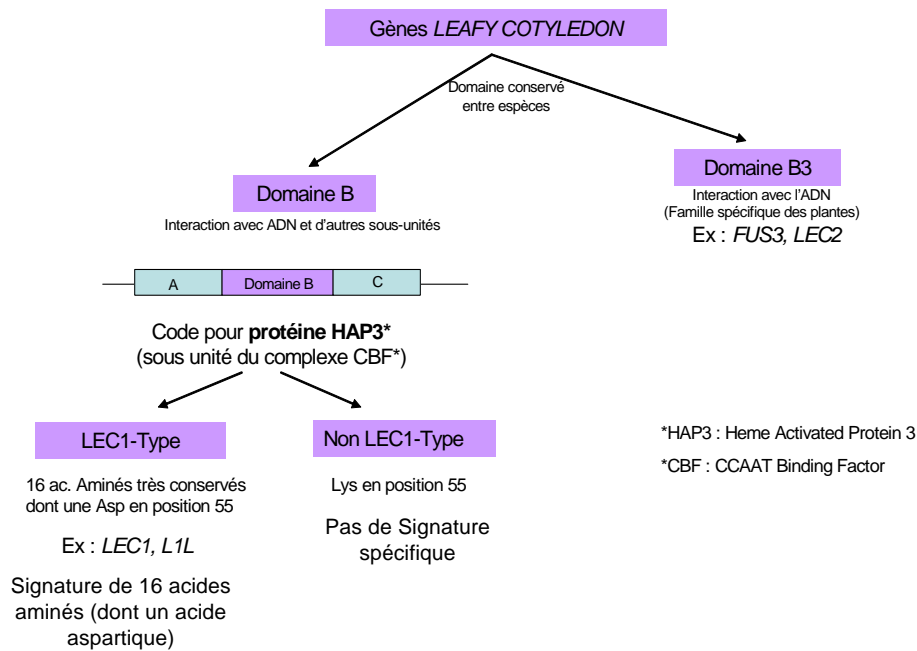
### **1. OBJECTIFS ET DEMARCHE**

Dans les deux premiers chapitres de ce travail de thèse, des approches transcriptomiques nous ont permis d'identifier un certain nombre de facteurs de transcription potentiellement impliqués dans le développement de la graine. Ce nouveau chapitre présente la caractérisation fonctionnelle de l'un des gènes identifiés précédemment.

Les gènes de type LEC1 ont été choisis comme gènes candidats pour l'élaboration de la composition protéique des graines en se basant sur les données disponibles chez *A. thaliana*. La question biologique de ce chapitre est de savoir : quels rôles jouent les gènes de type LEC1 lors du développement de la graine chez la légumineuse *M. truncatula*. Dans le temps imparti pour cette thèse, nous nous sommes focalisés sur l'un des deux gènes de type LEC1 identifiés chez *M. truncatula*, le gène *LEAFY COTYLEDONI-LIKE (MtLIL)*. Pour répondre à cette question, nous avons analysé l'effet de mutants "perte de fonction" pour ce gène.

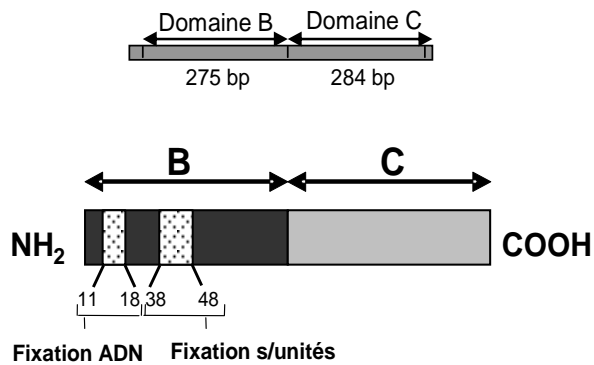
Ce chapitre est constitué d'un paragraphe introductif expliquant le choix des gènes de type LEC1 comme candidats pour le contrôle transcriptionnel du remplissage du grain suivi de l'identification et la caractérisation de deux gènes de type LEC1 chez *M. truncatula* (*MtLEC1* et *MtLIL*). Puis, l'implication du gène *MtLIL* dans le processus de développement de la graine est décrite grâce à des mutants "perte de fonction". Enfin, le rôle de ce gène lors du développement de la graine est discuté.



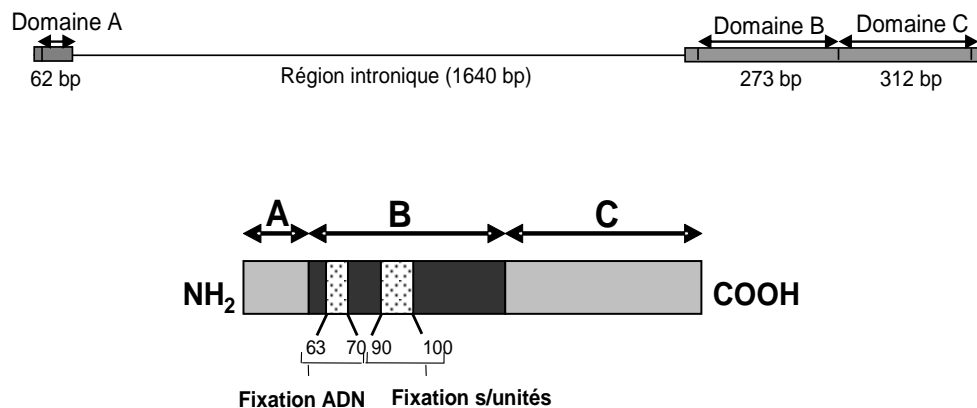


**Figure 3.1 : Description des gènes *LEAFY COTYLEDON* (d'après Kwong *et al.*, 2003)**

**Le gène *MtLEC1*(1347.m00026)**



**Le gène *MtL1L* (BI311277)**



**Figure 3.2 : Représentation des gènes de type LEC1 identifiés chez *M. truncatula*.**

Les séquences génomiques sont indiquées avec la taille et la position des domaines.

La séquence protéique est également indiquée avec la position des zones conservées de fixation à l'ADN et de fixation des sous-unités.

## 2. INTRODUCTION

### 2.1 Choix des gènes de type LEC1 comme candidats

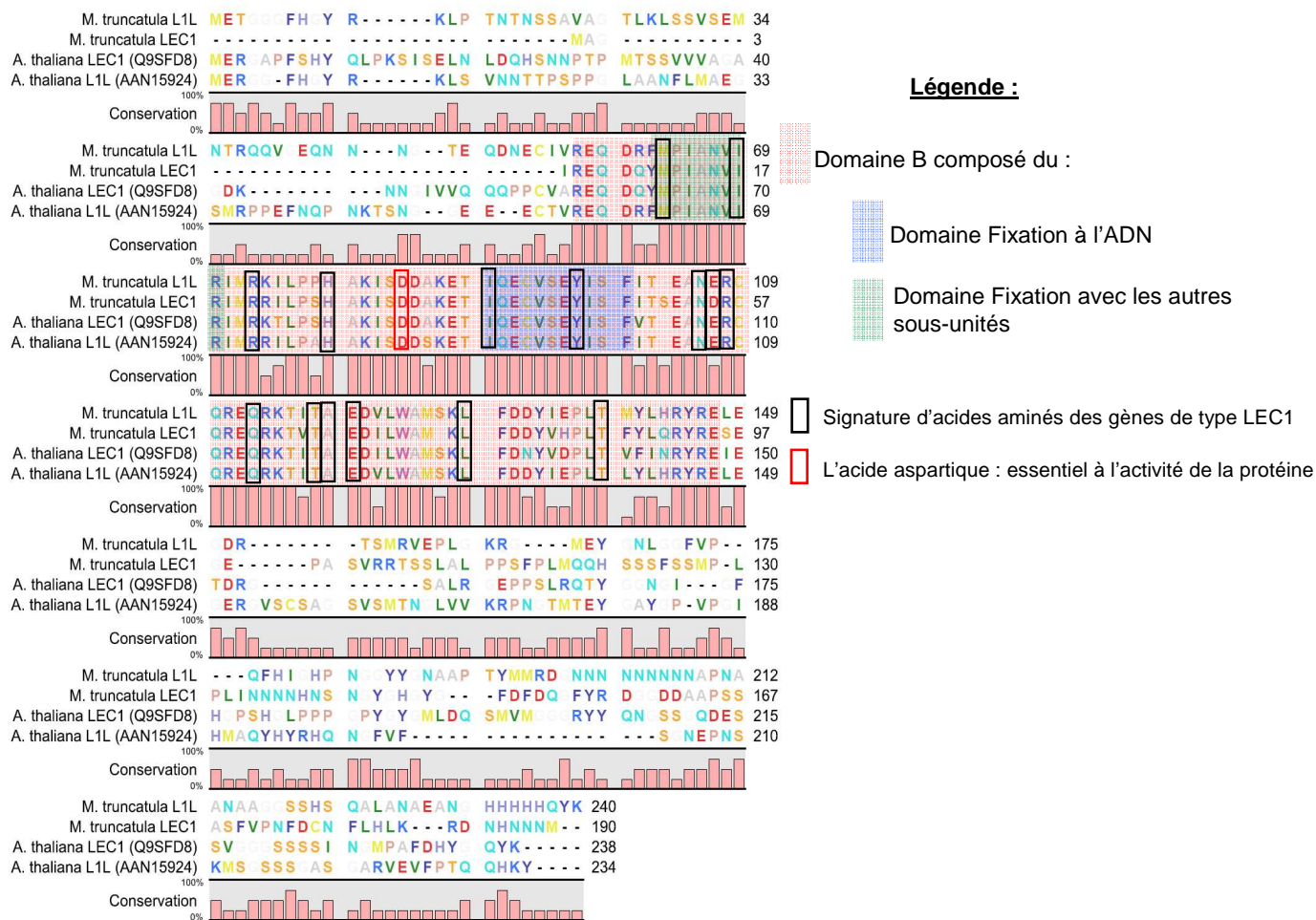
Pour des raisons liées au temps nécessaire de "purification" des fonds génétiques des plantes mutées et de régénération *in vitro* de plants transformés, il a fallu très vite choisir le(s) gène(s) à analyser afin de pouvoir obtenir des lignées mutantes "utilisables" dans le temps imparti pour la thèse. Ce choix s'est fait selon plusieurs critères : tout d'abord, nous voulions un gène exprimé de manière précoce dans le développement avec un pic d'expression au moment du déclenchement de la synthèse des protéines de réserve. Nous voulions également choisir un gène sur lequel des données étaient déjà disponibles dans la littérature afin de valider nos approches "perte de fonction".

A partir de la liste de facteurs de transcription candidats identifiées précédemment en PCR quantitative à haut-débit (Chapitre 2), un seul gène répondait à tous ces critères, le gène de type LEC1 (1347.m00026). En effet, ce gène a été identifié comme exprimé dans l'embryon et l'albumen au début de remplissage de la graine chez *M. truncatula*. Le rôle des gènes de type LEC1 dans le développement de la graine d'*A. thaliana* est bien documenté en ce qui concerne le gène *LEC1* (West *et al.* 1994, Kagaya *et al.* 2005a) et reste encore à définir pour les gènes *LEC1-LIKE* (Kwong *et al.* 2003). Et enfin, l'homologue d'un gène de type LEC1 a pu être identifié chez le pois (voir Chapitre 4).

### 2.2 Caractéristiques des gènes *LEAFY COTYLEDON* : cas particulier des type LEC1

Les gènes de type *LEAFY COTYLEDON* (*LEC*) sont connus pour être des gènes maîtres du développement de la graine en raison de leur rôle fondamental lors de ce processus. Ces gènes (*LEC1*, *LEC2* et *FUS3*) sont des régulateurs clefs du développement des graines (pour revue Harada, 2001). Les mutants associés à ces gènes montrent des phénotypes communs. Tout d'abord, la formation de l'embryon est anormale et présente une transformation des traits embryonnaires en caractéristiques de plantules : transformation des cotylédons en feuilles avec apparition de trichomes et présence d'un système vasculaire complexe (Meinke *et al.* 1994). Ensuite, les mutants *lec* présentent des défauts dans la synthèse et dans l'accumulation de protéines spécifiques des graines dont les protéines de réserve (Keith *et al.* 1994 ; Meinke *et al.* 1994). Enfin, les graines mutantes pour ces gènes présentent également des défauts de tolérance à la dessiccation à des degrés divers (Meinke 1992 ; Baumlein *et al.* 1994 ; Keith *et al.* 1994 ; Meinke *et al.* 1994 ; West *et al.* 1994).

Ces gènes *LEC* appartiennent à deux familles distinctes de facteurs de transcription en fonction de leur domaine de fixation à l'ADN : ceux possédant un domaine de type B3 (*LEC2* et *FUS3*) qui appartiennent à une famille de gènes spécifiques des plantes et ceux possédant un domaine de type B (Figure 3.1). Pour les gènes possédant ce domaine de fixation à l'ADN de type B, ils constituent la famille des HEME ACTIVATED PROTEIN3 (HAP3) qui constitue une sous-unité du CCAAT Binding Factor (CBF aussi connu par l'abréviation NF-Y chez la levure) (Lothan *et al.* 1998). Ce complexe CBF est très largement représenté chez les eucaryotes et permet



**Figure 3.3 :** Alignement des séquences protéiques de MtL1L (BI311277), MtLEC1 (1347.m00026) AtLEC1 (Q9SFD8) et AtL1L (AAN15924). La position des domaines conservés est indiquée ainsi que la signature d'acides aminés spécifiques des gènes de type LEC1. La conservation de séquence est représentée sous forme d'une barre rose indiquant une valeur de 0% à 100% d'identité.

% Identité (% homologie)	MtL1L (BI311277)	MtLEC1(1347.m00026)	AtLEC1 (Q9SFD8)	MtL1L(Domaine B)	MtLEC1(Domaine B)	AtLEC1 (Q9SFD8)(Domaine B)
MtL1L (BI311277)		34,9% (40,8%)				
MtLEC1(1347.m00026)	34,9% (40,8%)					
AtLEC1 (Q9SFD8)	48,2% (58,6%)	42,7% (50,4%)				
AtL1L (AAN15224)	52,3% (58,8%)	37,4% (41,2%)	41,2% (51,9%)			
MtL1L (Domaine B)					85,5% (93,4%)	
MtLEC1(Domaine B)				85,5% (93,4%)		
AtLEC1 (Q9SFD8)(Domaine B)				85,7% (97,8%)	85,7% (93,4%)	
AtL1L (AAN15224)(Domaine B)				95,6% (98,9%)	85,7% (94,5%)	83,5% (98,9%)

**Tableau 3.1 :** Alignements des séquences protéiques de MtLEC1 (1347.m00026) et MtL1L (BI311277) avec leurs homologues chez *A. thaliana*. Les résultats d'alignements sont donnés pour les séquences entières mais également pour les domaines B. Les pourcentages d'identités sont indiqués ainsi que les pourcentages d'homologies entre parenthèses.

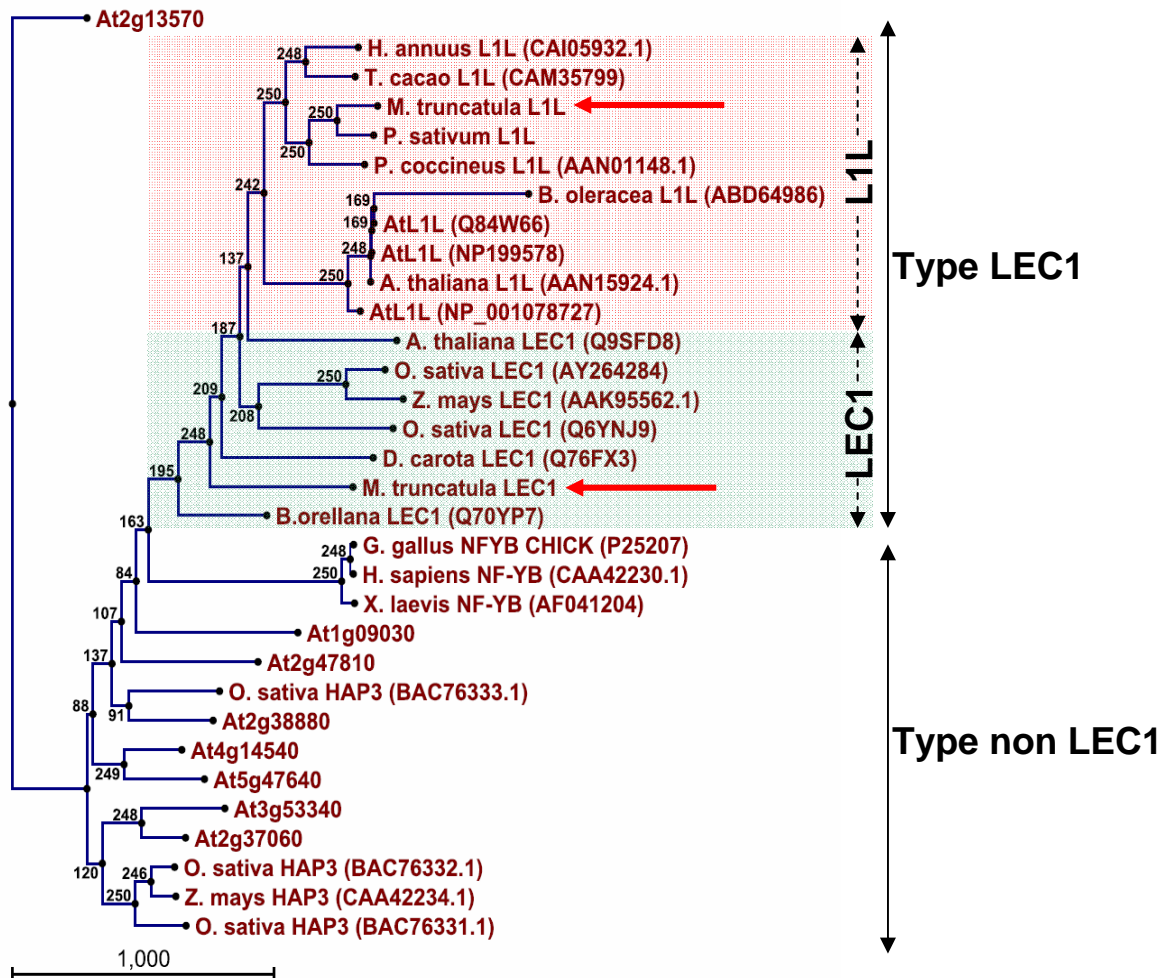
d'activer l'expression de nombreux gènes en se fixant à la boîte CCAAT présente dans les régions promotrices des gènes (Maity et De Crombrughe, 1998). Chez les plantes, ce complexe est hétérotrimérique et constitué des sous-unités HAP2, HAP3 et HAP5. A la différence de la levure, chez *A. thaliana* plusieurs formes de chaque sous-unité ont été identifiées, impliquant des rôles multiples dans la transcription des gènes (Edward *et al.* 1998). Plusieurs gènes de type HAP3 ont été identifiés et séparés en deux classes en fonction de la séquence d'acides aminés de leur domaine B. **La première classe contient le gène *LEC1* et les gènes *LEC1-LIKE (LIL)* et forme le groupe des gènes de type *LEC1*.** Les autres sous-unités HAP3 sont regroupées dans la seconde classe des gènes non-*LEC1* type. Les types *LEC1* diffèrent des non-*LEC1* par une signature protéique dans le domaine B, constituée de 16 acides aminés, très conservée dont un résidu acide aspartique (Asp, D), en position 55 chez *A. thaliana*, qui est essentiel à l'activité de la protéine (Lee *et al.* 2003). En revanche, les gènes de type non-*LEC1* ne possèdent pas de signature spécifique, à l'exception de la présence d'une lysine (Lys, K) en lieu et place de l'acide aspartique (Kwong *et al.* 2003).

### 3. RESULTATS

#### **3.1. Identification de deux gènes de type *LEC1* chez *M. truncatula***

L'analyse de PCR quantitative haut débit, présentée dans le chapitre précédent, nous a permis d'identifier un gène de type *LEC1* (1347.m00026). Ce gène, isolé à partir d'un clone BAC (AC51709\_26), correspond à une séquence codante de 573 nucléotides qui code une protéine de 191 acides aminés (Figure 3.2a). Le poids moléculaire théorique de cette protéine ainsi que son point isoélectrique sont respectivement de 21,6 kDa et 5,59. La protéine est constituée d'une séquence d'acides aminés proche de celle des gènes de type *LEC1*. Son domaine B se situe entre les acides aminés 13 et 96 et il présente de fortes homologies (plus de 93%) avec ceux présents dans les protéines de type *LEC1* chez *A. thaliana* (Tableau 3.1). Ce domaine est composé du site de fixation à l'ADN (11 MPIANVIR 18), ainsi que du site de fixation des sous-unités coactivatrices du CBF (38 IQECVSEYISF 48) (Figure 3.2a). Le domaine B est composé de 15 des 16 acides aminés qui constituent la signature des gènes de type *LEC1* dont l'acide aspartique (D) en position 32 essentiel à l'activité du domaine B (Figure 3.3). Le seul acide aminé qui diffère de la signature est un acide aspartique (D) qui est remplacé par un acide glutamique (E), acide aminé qui possède les mêmes propriétés physico-chimiques. L'analyse de la séquence génomique présente sur le clone BAC n'a pas permis de mettre en évidence de région intronique.

Nous avons ensuite utilisé la séquence protéique précédente, du gène 1347.m00026, ainsi que celles des gènes de type *LEC1* d'*A. thaliana* pour identifier d'autres séquences présentes dans les bases de données génomiques chez *M. truncatula*. Un alignement local en utilisant l'algorithme BLASTX (Basic Local Alignment Search Tool) a été réalisé entre les séquences protéiques des gènes de type *LEC1* et la base de ressources génomiques du Dana-Farber Cancer Institute *Medicago truncatula* gene index (DFCI MtGI version 8). Six séquences homologues à celles de gènes de type HAP3 ont été identifiées, une seule cependant, l'accession BI311277, présentait dans son domaine B la signature des gènes de type *LEC1*, les cinq autres correspondaient à



**Figure 3.4 : Arbre phénétique des gènes HAP3.**

Cet arbre a été généré grâce à l'algorithme Neighbor Joining (bootstrap de 500). Les accessions GeneBank pour toutes les séquences protéiques sont indiquées entre parenthèses. La classe des gènes de type LEC1 est indiquée, ainsi que les deux sous-classes correspondantes aux gènes *LEC1* (en vert) et *L1L* (en rouge). Les deux gènes de *M. truncatula* sont signalés par des flèches rouges.

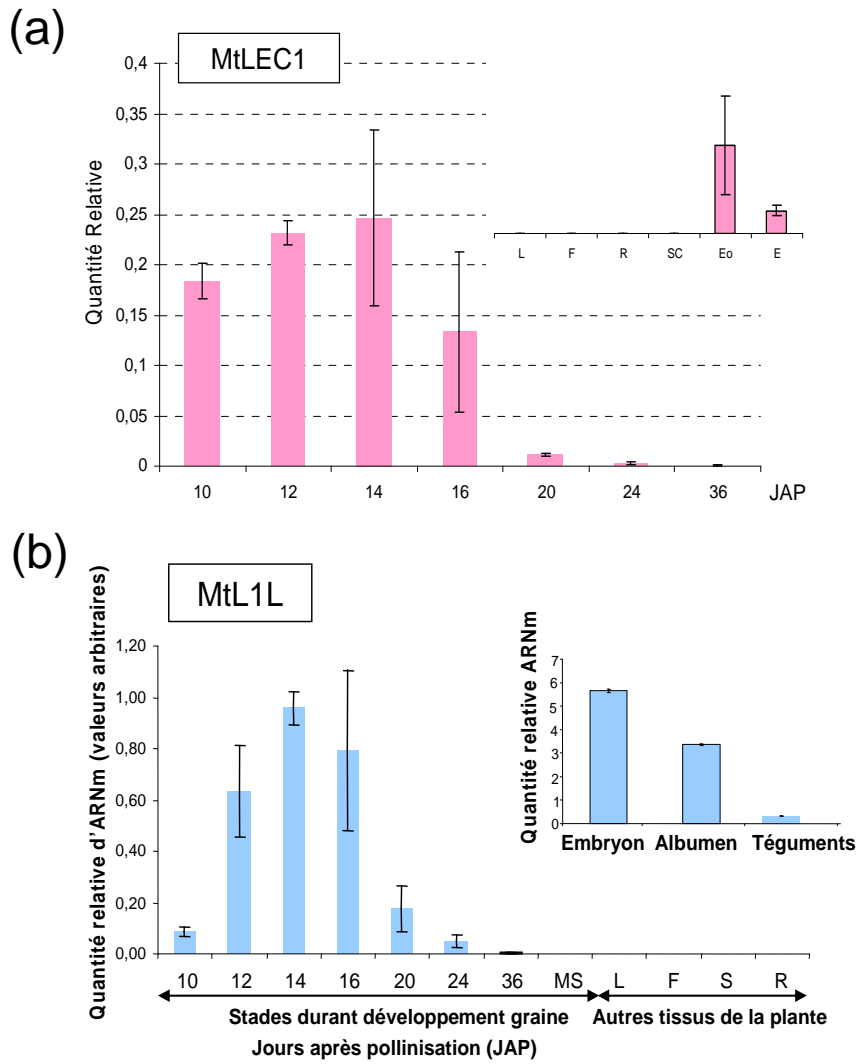
des HAP3 de type non-LEC1. Ce nouveau gène de type LEC1, isolé à partir d'une EST (BI311277), correspond à une région codante de 723 pb qui code pour une protéine de 241 acides aminés. Le poids moléculaire et le point isoélectrique théorique de cette protéine sont respectivement de 26,7 kDa et 6,26. La séquence protéique est homologue à celle d'*AtLEC1* et *AtLIL* (58% d'homologie, Tableau 3.1) et le domaine B présente une homologie de 98%. Ce domaine compris entre les acides aminés 56 et 147, est constitué des deux régions spécifiques : le site de fixation à l'ADN (63 MPIANVIR 70) et le site de fixation des sous-unités (90 IQECVSEYISF 100) (Figure 3.2b). La signature des 16 acides aminés dont l'acide aspartique, en position 84, est également présente (Figure 3.3). La séquence génomique de ce gène a ensuite été analysée ce qui a permis de mettre en évidence une région intronique de 1640pb entre les domaines A et B (Figure 3.2b).

### 3.1.1 Comparaison de séquences

Des alignements de séquences protéiques (avec l'algorithme NEEDLE (Needleman et Wunsch, 1970)) ont été réalisés entre les gènes d'*A. thaliana* (*AtLEC1*, Q9SFD8 et *AtLIL*, AAN15224) et les gènes identifiés chez *M. truncatula*. Ces résultats sont présentés dans le tableau 3.1, ils révèlent une très forte homologie entre les gènes identifiés chez *M. truncatula* et ceux codant les gènes de type LEC1 chez *A. thaliana*. On constate également que les domaines B de ces quatre gènes (*AtLEC1*, *AtLIL*, 1347.m00026 et BI311277) sont très proches avec une homologie supérieure à 85%. L'analyse des séquences protéiques complètes révèle que la protéine correspondante au gène 1347.m00026 présente plus d'homologie avec *AtLEC1* (50,4 %), tandis que celle du gène BI311277 est plus proche d'*AtLIL* (58,8 %).

Une analyse des séquences génomiques et protéiques des deux gènes identifiés chez *M. truncatula* montre une différence majeure. En effet, le gène BI311277 est composé de trois domaines A, B et C, avec un intron situé entre les domaines A et B au niveau de la séquence génomique. En revanche, le gène 1347.m00026 est plus court et ne possède que les domaines B et C, aucun exon correspondant au domaine A n'a pu être identifié par une recherche de séquences homologues dans le clone BAC contenant le gène (BAC AC151709) (Figure 3.2).

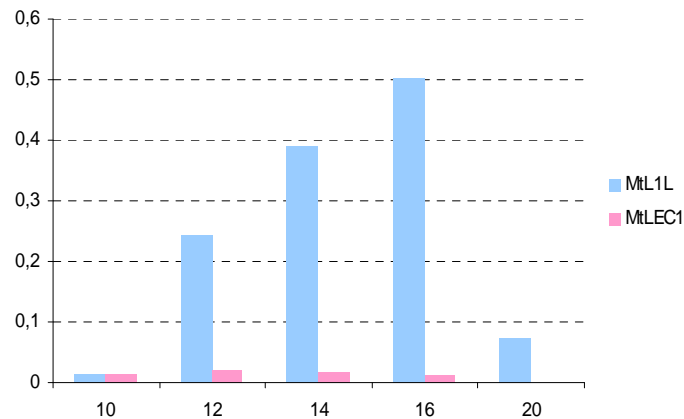
Une comparaison phénétique des gènes de *M. truncatula* a été réalisée en utilisant les séquences protéiques de plusieurs gènes HAP3 présents chez d'autres espèces végétales et animales. Dans cette arbre, nous retrouvons bien l'organisation décrite précédemment, à savoir la distinction entre les séquences protéiques HAP3 de type non-LEC1 et celle de type LEC1 (Figure 3.4). De plus, un autre niveau d'organisation en deux sous-classes apparaît à l'intérieur des gènes de type LEC1 : on constate une séparation entre ceux similaires à *LEC1* (*LEAFY COTYLEDON 1*) et ceux similaires à *LIL* (*LEAFY COTYLEDON 1-LIKE*) (figure 3.4). Cette observation confirme les résultats obtenus précédemment, à savoir que le premier gène identifié (1347.m00026) se rapproche plus des gènes *LEC1* alors que l'EST (BI311277) est plus proche des gènes *LIL*. Ainsi dans la suite de ce travail, ces gènes prendront respectivement le nom de *MiLEC1* (1347.m00026) et *MiLIL* (BI311277).



**Figure 3.5 : Profils d'expression des gènes (a) *MtLEC1* et (b) *MtL1L*.**

Profils obtenus en PCR quantitative en utilisant le gène de référence *MSC27*.

Les valeurs au cours du développement de la graine sont indiquées, ainsi que les tissus de la graine (embryon, albumen, Eo et téguments, SC). Les expressions dans les feuilles (L), les fleurs (F), les tiges (S) et les racines (R) sont également présentées.



**Figure 3.6 : Comparaison de l'accumulation des transcrits de *MtLEC1* et *MtL1L*.**

Profils réalisés en PCR quantitative en utilisant le gène de référence *MSC27*.



### 3.1.2 Analyse des profils d'expression

Afin d'analyser plus en détail ces gènes, nous avons regardé leur niveau d'expression au cours du développement de la graine et dans les différents organes de la plante.

Le profil d'expression en PCR quantitative du gène *MtLECI* avait déjà été réalisé puisqu'il avait été sélectionné en fonction de ce critère suite à l'analyse de PCR quantitative à haut débit (Chapitre 2). Ce profil a cependant été confirmé en utilisant des ADNc provenant d'une troisième répétition biologique. Ce gène présente donc un profil d'expression spécifique à la graine avec un maximum d'accumulation des ARNm aux stades 12-14 JAP, correspondant à la transition embryogenèse-remplissage et au début du remplissage. Ensuite, son expression diminue jusqu'à devenir quasiment nulle à partir de 20 JAP et disparaître à 24 JAP. Dans les tissus de la graine, nous avons confirmé l'expression de *MtLECI* au niveau de l'albumen et de l'embryon au stade 14 JAP (Figure 3.5a).

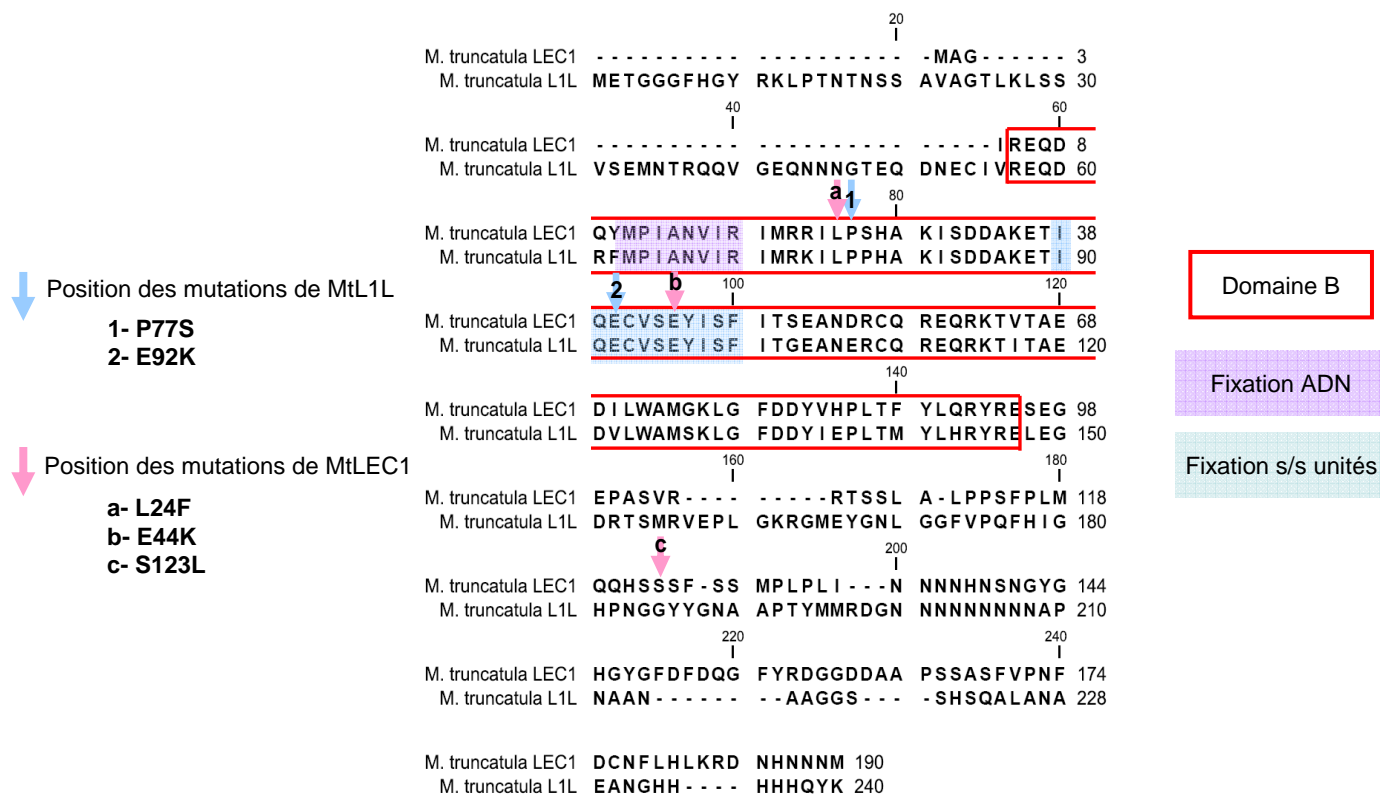
Quant au gène *MtLIL*, nous disposions dans un premier temps des données du northern électronique de l'EST, à savoir que cette séquence avait été identifiée dans une banque d'ADNc réalisée à partir de graines en début de développement. Une analyse d'expression en PCR quantitative a confirmé ce résultat et a montré une expression du gène spécifique à la graine avec une accumulation croissante de 10 à 14 JAP pour atteindre un maximum d'expression à 14-16 JAP, pour finalement diminuer très fortement à 20-24 JAP (Figure 3.5b).

Les gènes *MtLECI* et *MtLIL* présentent donc des expressions chevauchantes caractérisés par un léger décalage de leur maximum d'expression. Une comparaison des niveaux d'accumulation de ces deux transcrits a ensuite été réalisée à partir des mêmes échantillons d'ADNc et dans les mêmes conditions de PCRq (Figure 3.6). De manière intéressante, une grande différence dans les niveaux d'accumulation des transcrits a été observée. Les transcrits de *MtLIL* s'accumulent environ 25 fois plus que ceux de *MtLECI*.

### **3.2 Recherche de mutants "perte de fonction" pour ces gènes chez *M. truncatula***

Dans le but de réaliser une caractérisation fonctionnelle de ces deux gènes chez *M. truncatula*, nous avons recherché des mutants pour ces gènes en utilisant la méthode de criblage TILLING sur les populations de mutants EMS présentes au laboratoire. En parallèle, une méthode de transgenèse a été utilisée pour inactiver le gène *MtLIL* par RNA interférence (RNAi). Les banques de mutants Tnt1 (Gif/Yvette et Noble Foundation) ainsi que les populations Fast Neutron (John Innes Center et Noble Foundation) ont également été criblées mais aucun mutant pour ces gènes n'a pu être identifié dans ces populations.





**Figure 3.7 : Séquences protéiques des gènes *MtL1L* et *MtLEC1* et position des mutations identifiées par criblage TILLING.**

Le domaine B et les deux régions essentielles du domaine sont indiqués, ainsi que la position et le nom des différentes mutations.

Gènes	Identifiant	Population Mutants	ID	Type de mutation	Propriétés acides aminés	SIFT Prediction
<i>MtLEC1</i>	1347.m00026	EMS (UMRLEG)	AC151709_26	E 44 K S 123 L L 24 F	Acide à Basique Hydrophile à Hydrophobe même propriété	Affecte la fonction Affecte la fonction Affecte la fonction
<i>MtL1L</i>	BI311277	EMS (UMRLEG)	BI311277	P 77 S E 92 K	Hydrophobe à Hydrophile Acide à Basique	Affecte la fonction Affecte la fonction

**Tableau 3.2 : Identification des mutations TILLING.**

Le nom des gènes, l'identifiant GeneBank sont indiqués avec la nature des mutations et leur position sur la séquence protéique. Le changement de propriétés physico-chimiques de l'acide aminé substitué et la prédiction faite par SIFT figurent en face des mutations correspondantes.

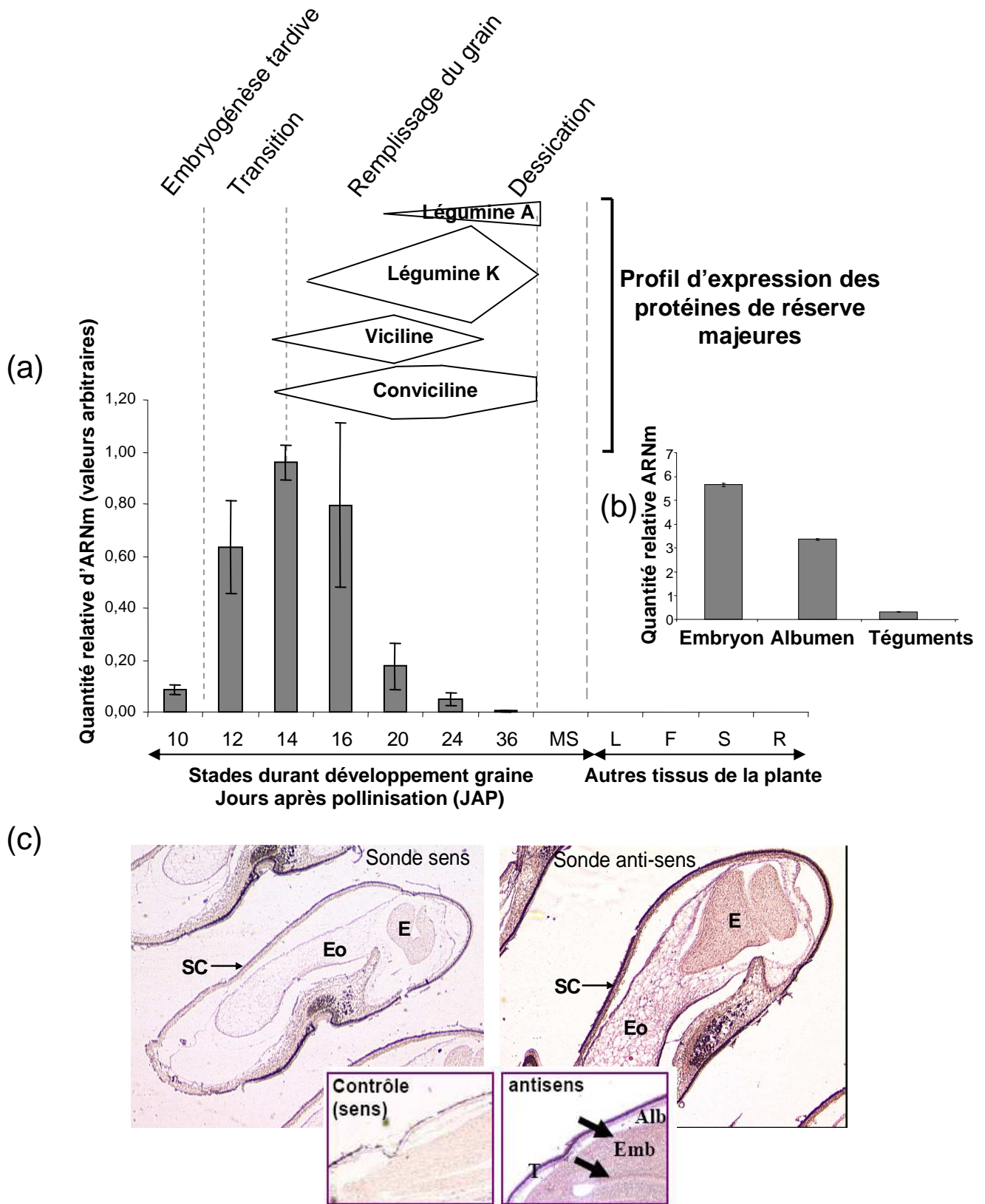
### 3.2.1 Criblage TILLING

Afin d'identifier des plantes mutées pour les deux gènes identifiés chez *M. truncatula*, les populations de mutants EMS présentes au laboratoire ont été criblées en utilisant la méthode TILLING (voir protocole criblage TILLING et Figure 2.3 du chapitre 2).

Pour *MtLECI*, trois plantes contenant trois mutations différentes dans la séquence codante du gène ont été identifiées. La première plante porte une mutation correspondant au changement en position 44 de la protéine d'un résidu acide glutamique (E) en lysine (K) (mutant E44K), changement d'acide aminé acide en basique. Cette modification est très intéressante car elle est située dans le domaine B et plus précisément dans la région très conservée qui est utilisée dans la fixation des sous-unités activatrices du CBF. La deuxième plante porte une mutation en position 123, ce qui correspond au changement d'un résidu sérine (S) en leucine (L) (mutant S123L), changement d'un acide aminé hydrophile en hydrophobe. Cette mutation, comme la précédente, correspond à une substitution d'acide aminé qui entraîne une modification physico-chimique pouvant modifier la conformation de la protéine et ainsi affecter sa fonction. La troisième plante porte une mutation qui correspond au changement d'une lysine (L) en une phénylalanine (F) en position 24 de la protéine (mutant L24F), cette mutation bien que située dans un domaine très conservé paraît moins intéressante que les deux précédentes (Figure 3.7). En effet, l'acide aminé est substitué par un autre présentant les mêmes propriétés, donc cette mutation est peu susceptible d'affecter la fonction de la protéine. Une analyse prédictive sur l'effet d'un SNP dans une séquence donnée (algorithme SIFT, <http://blocks.fhcrc.org/sift/SIFT.html>) suggère que les trois mutations identifiées pourraient affecter la fonction de la protéine, y compris la troisième où la substitution n'affecte pas les propriétés physico-chimiques de l'acide aminé. Nous avons donc gardé ces trois mutations pour l'analyse (Tableau 3.2).

Pour *MtLIL*, deux plantes ont été identifiées avec des mutations situées dans le domaine conservé B. La première plante mutée présente un changement d'une proline (P) en sérine (S), changement d'un résidu hydrophobe à hydrophile en position 77 de la protéine (mutant P77S). La deuxième mutation correspond à un changement d'un acide glutamique (E) en lysine (K), substitution d'un acide aminé acide en basique. Ce changement est intéressant car situé en position 92 de la protéine (mutant E92K) dans la région de fixation des sous-unités du complexe CBF (Figure 3.7). Les prédictions SIFT suggèrent que ces deux mutations pourraient affecter la fonction de la protéine *MtLIL* (Tableau 3.2).

Les cinq plantes portant ces mutations (3 pour *MtLECI* et 2 pour *MtLIL*) ont subi deux rétrocroisements avec des plantes sauvages afin d'éliminer une grande partie des mutations "accessoires" présentes dans le génome suite au traitement mutagène EMS.



**Figure 3.8 : Profils d'expression de Mtl1L pendant le développement de la graine et dans les différents tissus de la plante (L, feuilles ; F, fleurs ; S, tiges et R, racines).**

(a) et (b) Profils réalisés par PCR quantitative en utilisant le gène de référence MSC27. Les gènes utilisés pour représenter les protéines de réserve sont conviciline (TC100299), viciline (TC100302), légumine K (TC100253) et légumine A (TC100252)

(c) Localisation des ARNm de *MtL1L* dans la graine à 14 JAP par *hybridation in situ* (réalisée avec Samia Djennane).

### 3.2.2 Transformants III RNAi

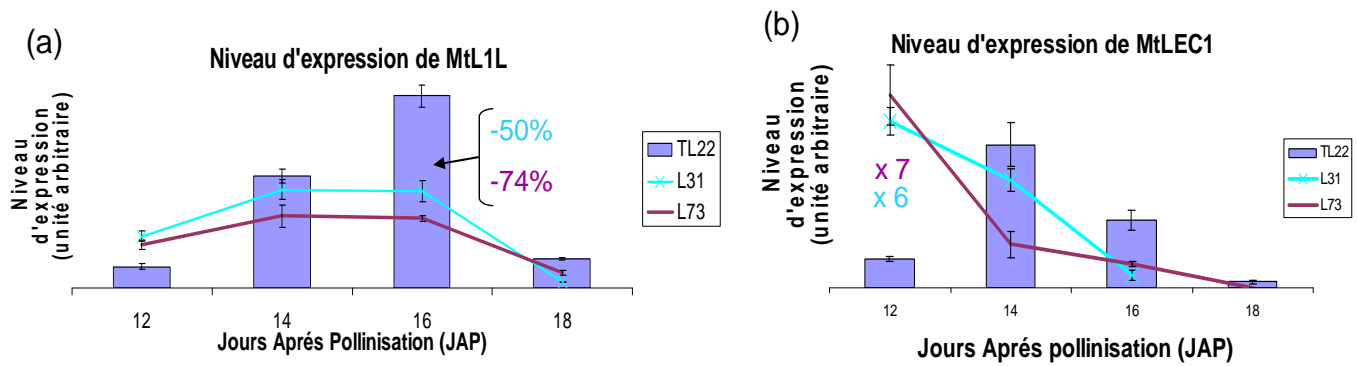
En parallèle à l'identification de mutants par TILLING, nous avons généré des plantes transformées présentant une inhibition de l'expression du gène *MtLIL* par RNA interférence. Pour cette approche, nous avons focalisé sur le gène *MtLIL* qui présentait un niveau d'expression plus important que celui de *MtLECI* et ainsi qui pourrait être moins sensible à d'éventuels phénomènes de complémentation par un autre gène endogène. Nous avons réalisé une construction de type "épingle à cheveu" avec un fragment du gène *MtLIL*, comportant une petite partie du domaine B et le domaine C (spécifique à chacun des gènes de type LEC1). Ce fragment du gène *MtLIL* a été cloné dans le plasmide pHELLGATE8 en orientation sens et antisens sous le contrôle du promoteur fort 35S du virus de la mosaïque du chou-fleur (Construction et régénération en Annexe 2) (Helliwell et Waterhouse 2003). Cette construction a ensuite été introduite chez *Agrobacterium tumefaciens* capable d'infecter des explants de feuilles de *M. truncatula*. Contrairement à tous les autres travaux de cette thèse, cette transformation a été réalisée sur l'écotype R108 de *M. truncatula* qui se prête mieux à la régénération cellulaire et au développement de cals sur milieux solides. Des embryons somatiques ont été obtenus puis les premières plantes contenant le transgène. Des analyses ont été réalisées sur ces plantes régénérées *in vitro* afin de vérifier l'intégration du transgène dans le génome. Ce transgène a été retrouvé chez deux plantes transformées III issues de deux cals différents (les plantes L73 et L31). Des plantes témoins sauvages R108 (plantes R108) et d'autres régénérées *in vitro* dans les mêmes conditions mais sans avoir subi la transformation (plantes TL22), nous ont servi de plantes contrôles.

### 3.3 Analyse fonctionnelle du gène *MtLIL*

Dans le temps imparti pour la thèse, nous avons choisi de nous concentrer sur l'analyse fonctionnelle du gène *MtLIL* qui est exprimé dans la graine à un niveau plus élevé que *MtLECI*. Par ailleurs, le rôle du gène *LIL* dans les différentes étapes du développement de la graine était moins documenté que *LECI* et nécessitait d'être plus finement élucidé.

Pour obtenir une première indication sur le rôle possible du gène *MtLIL*, nous avons comparé son profil d'expression à celui de gènes correspondant aux principales protéines de réserve (TC100299, TC100302, TC100253 et TC100252 codant respectivement pour une conviciline, viciline, légumine K et légumine A) en PCR quantitative. Le maximum d'accumulation des transcrits codant pour *MtLIL* coïncidait avec le début de synthèse des ARNm de convicilines et de vicilines et précédait celui des légumine K (Figure 3.8a). Une analyse d'expression du gène dans les différents tissus de la graine à 14 JAP a montré la présence d'ARNm de *MtLIL* dans l'embryon et dans l'albumen (Figure 3.8b).

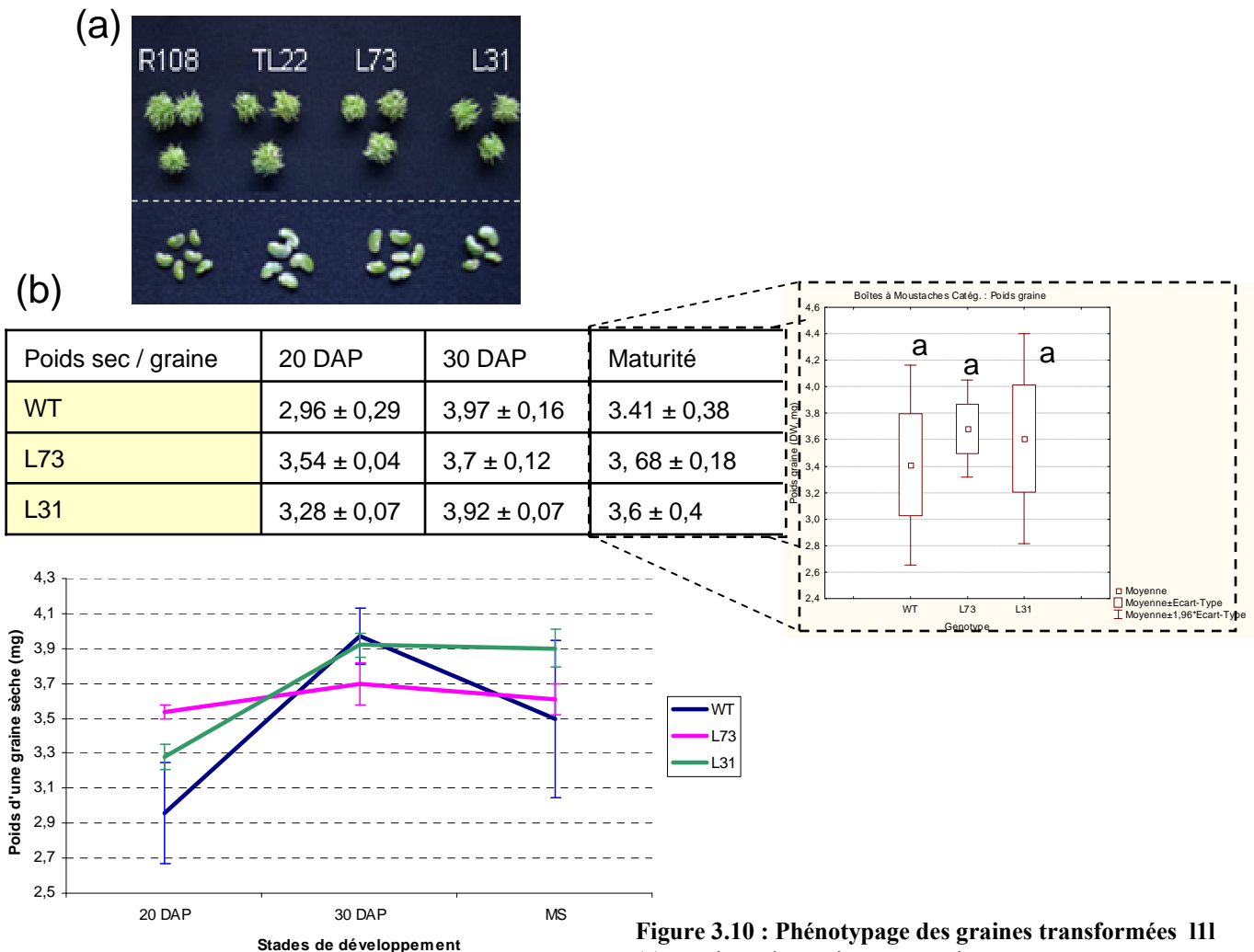
Une analyse plus fine de la localisation des transcrits à l'intérieur de la graine a également été entreprise par *hybridation in situ*. Cette analyse confirme les observations précédentes, à savoir que des transcrits de *MtLIL* sont présents essentiellement dans l'embryon à 14 JAP. En revanche, un signal d'hybridation aspécifique a été observé dans la graine entière et surtout dans le tégument, aussi bien dans les lames antisens que dans les lames



**Figure 3.9 : Analyse de l'expression de (a) *MtL1L* et (b) *MtLEC1* dans les plantes transformées III.**

Profils réalisés en PCR quantitative en utilisant le gène de référence *MSC27*. Le profil d'expression des plantes contrôles est indiqué par une barre alors que ceux des transformants sont indiqués par des lignes (bleue pour L31 et rouge pour L73).

L31 et L73 correspondent à deux plantes transformées indépendantes, TL22 correspond aux plantes témoins qui ont subies la régénération in vitro sans transformation.



**Figure 3.10 : Phénotypage des graines transformées III**

- (a) Photo des graines à maturité
- (b) Comparaison des poids de graines à 20 JAP, 30 JAP et à maturité entre les sauvages (R108 et TL22) et les transformants

contrôles sens, ce qui nous a empêché de conclure sur une localisation plus précise de ces transcrits dans la graine ou même à un niveau cellulaire.

### 3.3.1 Analyse des plantes transformées III

#### 3.3.1.1 Analyse de l'expression du gène *MtLIL*

Pour valider cette approche, nous avons dans un premier temps vérifié et quantifié la baisse d'expression du gène dans les plantes portant la construction RNAi. Le suivi l'expression du gène a été réalisé au cours des stades 12, 14, 16 et 18 JAP dans les plantes transformés L73 et L31 ainsi que dans les plantes témoins TL22. Les résultats obtenus, présentés dans la figure 3.9a, montrent qu'une baisse d'expression des transcrits est observée dans les plantes qui ont intégré l'insert. Cette baisse est marquée aux stades où le gène s'exprime fortement (16 JAP) et diffère entre les deux transformants. On constate que l'extinction d'expression est plus marquée pour les plantes L73 avec une baisse d'expression des transcrits de *MtLIL* d'approximativement 74 % à 16 JAP par rapport aux témoins (Figure 3.9a).

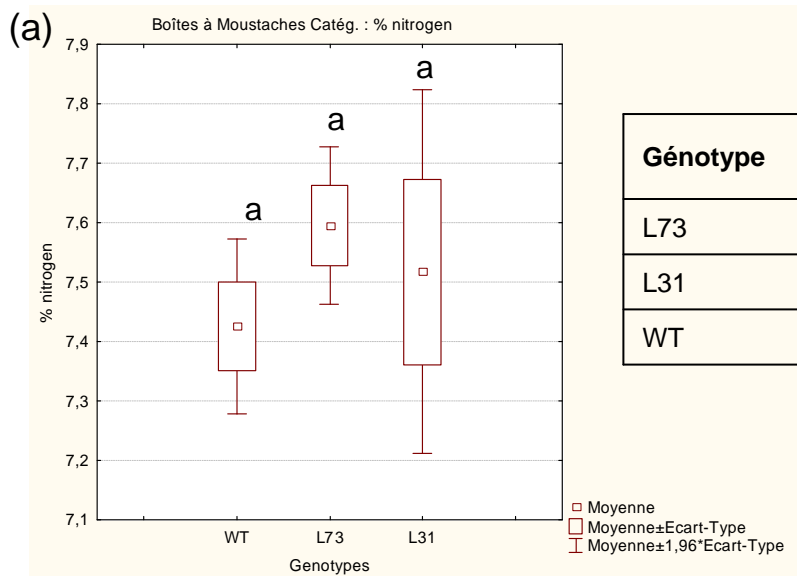
De manière intéressante, une analyse du niveau d'expression de *MtLECI* a également été réalisée chez ces mêmes plantes transformées. Même si cette expérience n'a été réalisée que sur une seule répétition biologique, on constate que l'expression de *MtLECI* est dérégulée chez les plantes transformées en comparaison aux témoins TL22 (Figure 3.9b).

#### 3.3.1.2 Phénotypage des plantes transformées RNAi

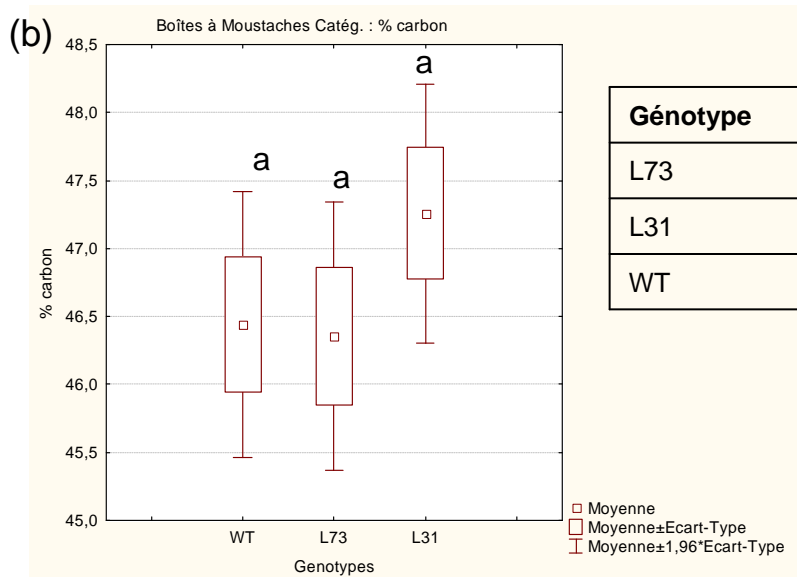
Le phénotypage des plantes transformées de *M. truncatula* a été réalisé sur la plante entière. Des caractères tels que le taux de germination, la taille des plantes, la précocité de floraison, le nombre de gousses par plante ont également été testés. Aucune différence n'a pu être mise en évidence entre les plantes transformées et les plantes témoins. Des cinétiques de remplissage des graines ont été réalisées et aucune différence sur le développement des graines n'a été mise en évidence (Figure 3.10a et 3.10b). Des analyses de teneur en carbone et azote (analyses Dumas, Allen 1974) ont également été réalisées et ont révélé une légère augmentation de l'azote des graines matures chez les plantes transformées L73 (Figure 3.11b). Cette tendance n'est pas statistiquement vérifiée ( $p$ -value $<0,08$ ) en raison d'un faible effectif de plantes testées et de la disparité des valeurs au sein des répétitions biologiques. Une mesure de la quantité de protéines dans les graines matures a également été entreprise par dosage Bradford. Ce dosage, tout comme la méthode Dumas, souligne une légère augmentation du contenu en protéines dans les graines des plantes transformées, augmentation qui n'est pas validée statistiquement (Figure 3.11c).

#### 3.3.1.3 Analyses biochimiques

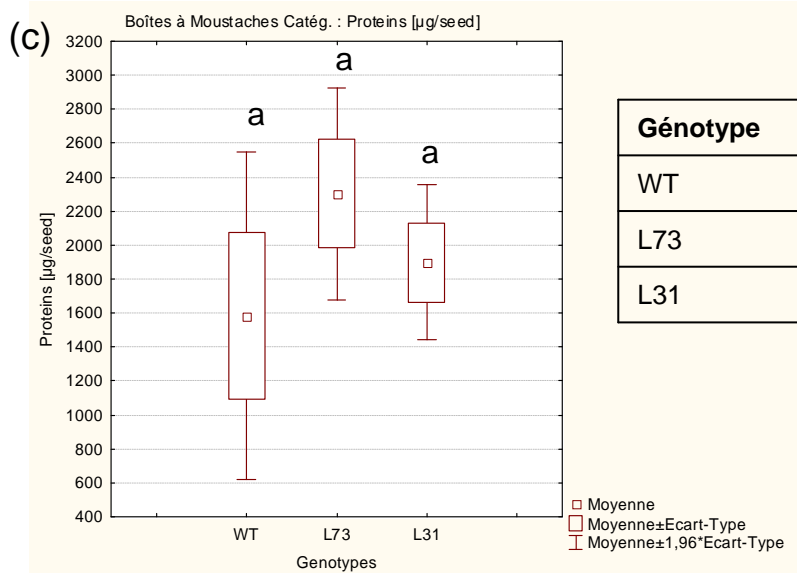
Pour confirmer cette accumulation différentielle d'azote et donc pour déterminer si l'extinction du gène *MtLIL* a un effet sur la composition protéique de la graine mature de *M. truncatula*, les extraits protéiques totaux des



Génotype	% nitrogène	N	Ecart-type
L73	7,594402	2	0,067503
L31	7,516805	2	0,156067
WT	7,425864	3	0,074732



Génotype	% carbone	N	Ecart-type
L73	46,35578	2	0,503008
L31	47,25787	2	0,484609
WT	46,44416	3	0,498592

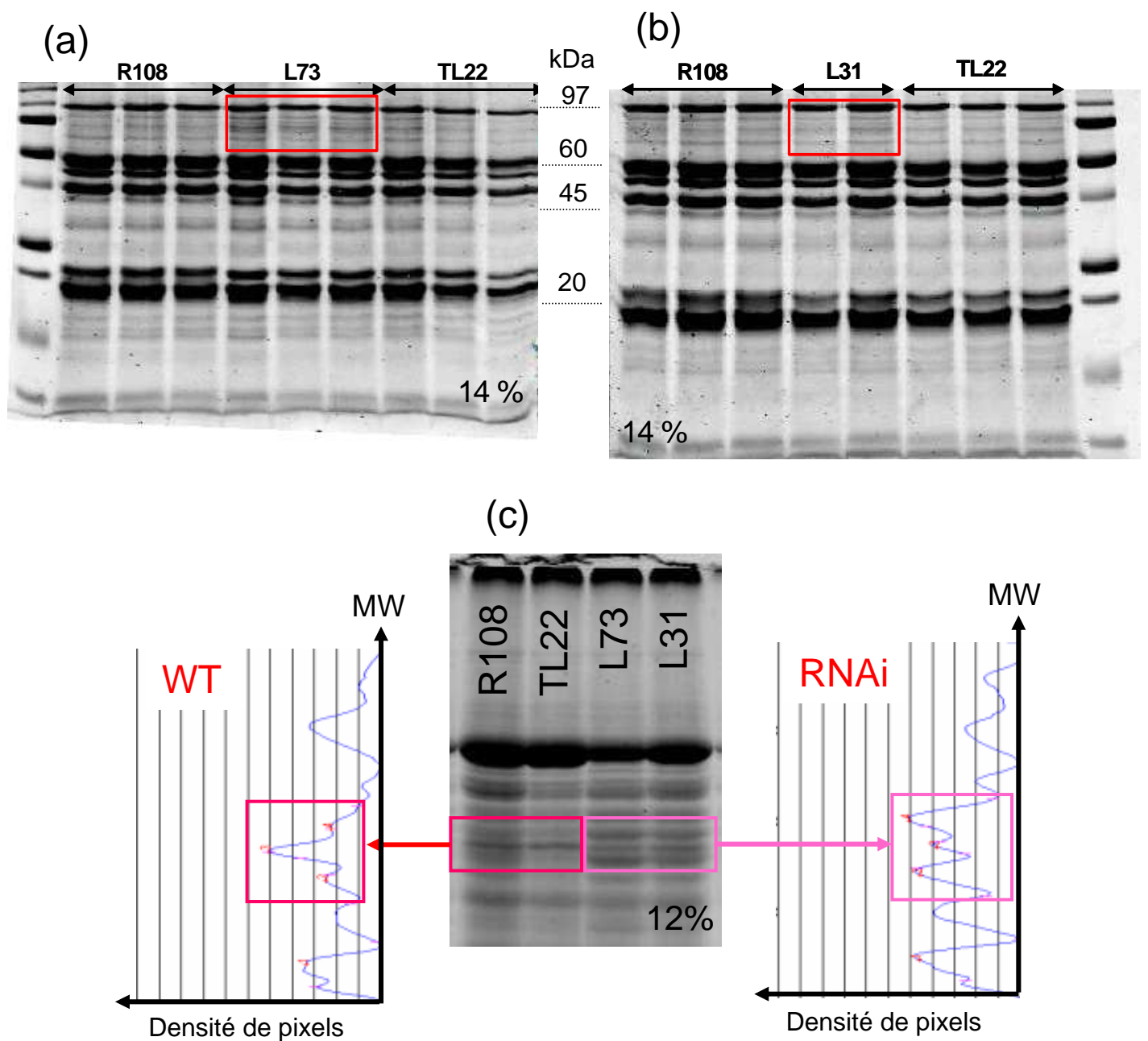


Génotype	Protéines [µg/seed]	N	Ecart-type
WT	1583,650	6	493,3296
L73	2302,318	3	318,5291
L31	1898,727	2	233,5672

**Figure 3.11 : Analyse de composition des graines matures transformées**

- (a) Analyse du contenu en azote des graines
- (b) Analyse du contenu en carbone des graines
- (c) Analyse de la concentration de protéines totales dans les graines (dosage Bradford)



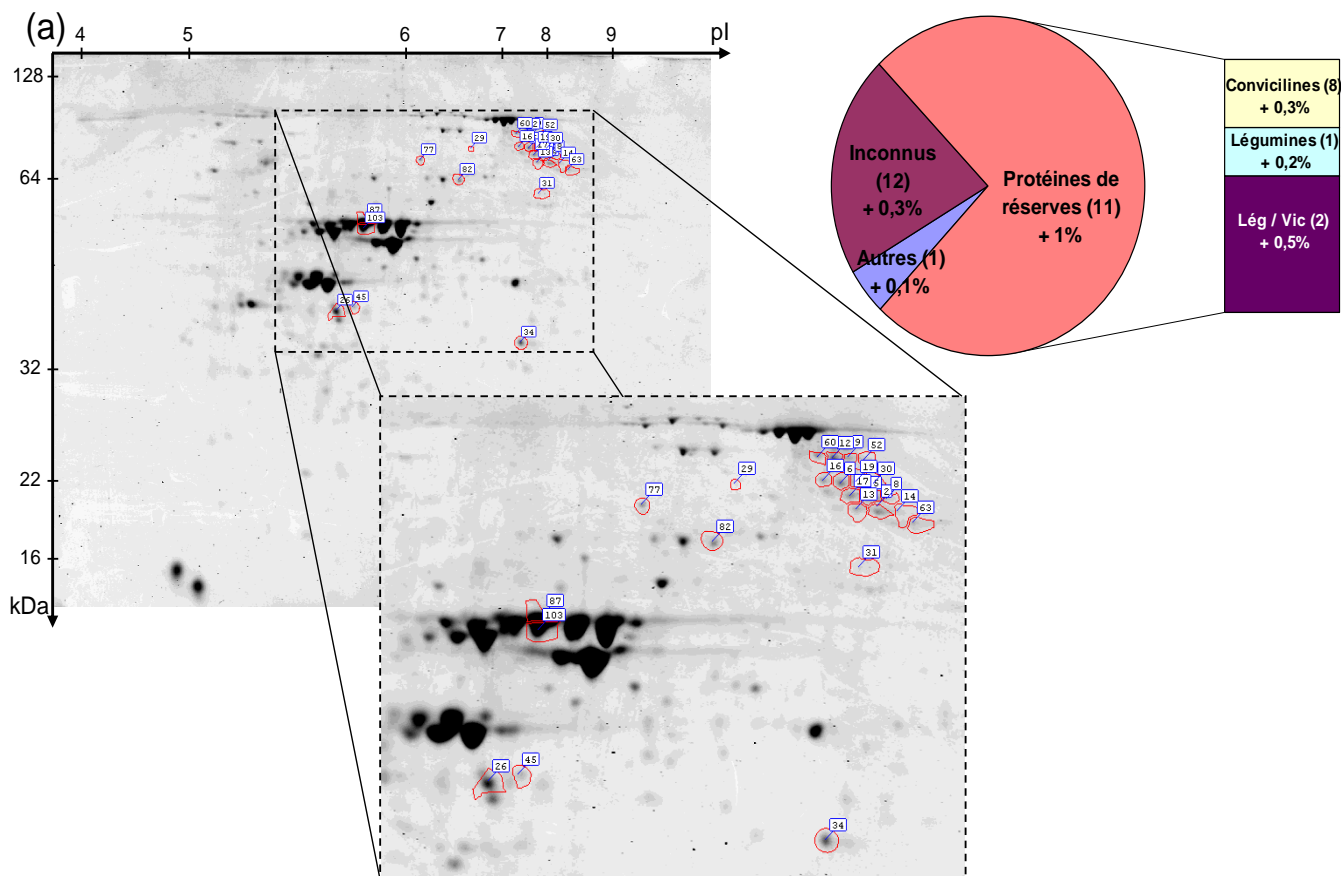


**Figure 3.12 : Composition de la graine mature transformées III**

(a) et (b) Variations de composition entre les graines matures sauvages (R108 et TL22) et (a) transformées L73 (a) et L31 (b). Chaque puits correspond à une répétition biologique.

(c) Un agrandissement de la zone permet de voir les différences sur gel à 12% d'acrylamide. Une analyse densitométrique en utilisant le logiciel Odyssey confirme ces observations.





**Figure 3.13 : Spots protéiques qui varient significativement dans les graines transformées III**

- (a) Gel présentant les numéros de spots qui varient dans les graines transformées et leur position  
 (b) Représentation des variations quantitatives par rapport au volume total du gel en fonction du type de protéines. Le nombre de spots concernés est indiqué entre parenthèses

N° Spot	Nom de la protéine	Numero d'accession	Classes fonctionnelles (MapMan)	Poids moléculaire et pI Théorique kDa / pI	Outil MS	Variation (par rapport au sauvage)	Abondance relative des spots (% du volume total)	
							WT	III
2	Unknown					2,4	0,054 +/- 0,014	0,128 +/- 0,076
5	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	2,1	0,061 +/- 0,014	0,126 +/- 0,049
6	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	2,0	0,041 +/- 0,014	0,085 +/- 0,036
8	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	2,0	0,019 +/- 0,004	0,037 +/- 0,022
9	Unknown					1,9	0,021 +/- 0,006	0,041 +/- 0,014
12	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	1,8	0,052 +/- 0,023	0,093 +/- 0,026
13	Unknown					1,7	0,053 +/- 0,016	0,09 +/- 0,038
14	Unknown					1,7	0,045 +/- 0,012	0,077 +/- 0,029
16	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	1,7	0,035 +/- 0,011	0,057 +/- 0,02
17	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	1,6	0,061 +/- 0,017	0,099 +/- 0,037
19	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	1,6	0,041 +/- 0,012	0,067 +/- 0,023
26	legumin A2	gi 22008	Development, Storage and LEA proteins		nanoESI Q-TOF	1,6	0,329 +/- 0,11	0,52 +/- 0,158
29	Unknown					1,5	0,005 +/- 0,002	0,007 +/- 0,002
30	Convicilin precursor	P13915	Development, Storage and LEA proteins	63,9 / 6,11	nanoESI Q-TOF	1,5	0,032 +/- 0,007	0,047 +/- 0,016
31	Unknown					1,5	0,044 +/- 0,011	0,067 +/- 0,02
34	Seed maturation protein PM34	Q9LLQ6	Development, Storage and LEA proteins	31,7 / 6,6	nanoESI Q-TOF	1,5	0,124 +/- 0,024	0,185 +/- 0,059
45	Unknown					1,5	0,051 +/- 0,009	0,077 +/- 0,015
52	Unknown					1,4	0,026 +/- 0,004	0,037 +/- 0,009
60	Unknown					1,3	0,037 +/- 0,011	0,05 +/- 0,013
63	Unknown					1,4	0,043 +/- 0,008	0,059 +/- 0,011
77	Unknown					1,3	0,034 +/- 0,007	0,043 +/- 0,012
82	Unknown					1,3	0,075 +/- 0,013	0,095 +/- 0,027
87	Vicilin / Legumin K	P13918 / S2668	Development, Storage and LEA proteins	49,3/5,33 et 56,5/5,65	nanoESI Q-TOF	1,4	0,509 +/- 0,068	0,712 +/- 0,138
103	Vicilin / Legumin K	P13918 / S2668	Development, Storage and LEA proteins	49,3/5,33 et 56,5/5,65	nanoESI Q-TOF	1,3	0,992 +/- 0,132	1,317 +/- 0,251

**Tableau 3.3 : Spots identifiés comme différentiellement exprimés dans les graines matures transformées III.**

Les numéros de spots se réfèrent à la figure 3.13

échantillons ont été analysés en gel dénaturant 1D. On a ainsi constaté une accumulation différentielle de bandes de haut poids moléculaire dans la région comprise entre 60 et 97 kDa, correspondant à la zone des convicilines (approx. 95 kDa) et des vicilines (approx. 60 kDa). Une analyse densitométrique de cette zone des gels 1D confirme ce résultat visuellement observé chez les deux transformants L73 et L31 (Figure 3.12).

Dans le but d'affiner l'analyse du contenu protéique des graines, les protéines totales des graines matures transformées et sauvages ont été séparées par électrophorèse bi-dimensionnelle (2D SDS PAGE). Les gels ont été colorés au bleu de Coomassie puis numérisés en utilisant un scanner Odyssey (LiCor) pour détecter et quantifier les spots. Puis, une analyse statistique (ANOVA à un facteur) a permis d'identifier les spots différentiellement exprimés entre les gels des graines matures témoins (R108 et TL22) et ceux des graines transformées (L73 et L31).

Les 24 spots protéiques qui varient significativement ( $p$ -value < 5% et puissance > 75%) sont répertoriés dans le tableau 3.3 et sur la figure 3.13. Les spots qui varient présentent tous une accumulation plus élevée chez les graines transformées l11, augmentation allant de 1,3 à 2,4. L'identification des spots a été réalisée par séquençage des polypeptides en LC-nano ESI-Q-TOF (Coll. INRA Tours, V. Labas). La moitié des polypeptides ont ainsi pu être identifiés et correspondent essentiellement à des protéines de réserve : 8 spots de convicilines (spots #5, 6, 8,12, 16, 17, 19, 30), 1 spot de légumine A2 (spot #26), 2 spots constitués d'un mélange de légumine K et de vicilines (spots # 87, 103). L'augmentation de l'abondance des 11 spots correspondant aux protéines de réserve expliquent environ 75% de la variation totale (Figure 3.13). De plus, parmi ces spots, nous pouvons constater que huit spots correspondent à des convicilines ayant subi des modifications post-traductionnelles. Cependant malgré le nombre de spots important de convicilines affectés dans ces graines transformées, l'augmentation du volume total est du à plus de 50% aux spots #87 et #103, identifiés comme des spots de légumine K et vicilines en mélange.

### 3.3.2 Analyse des mutants EMS/TILLING *MtLIL*

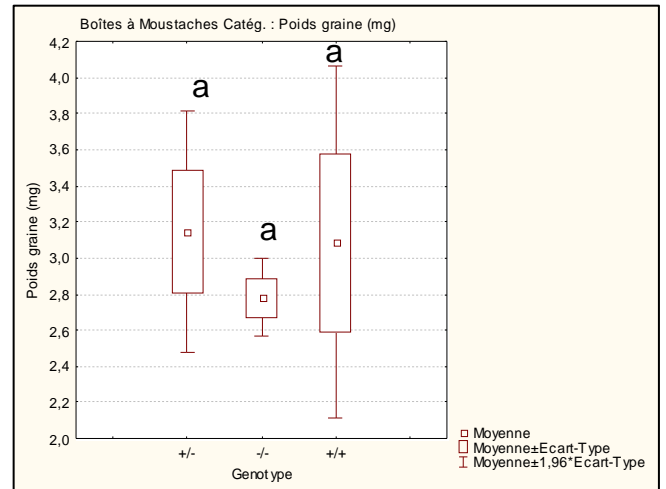
#### 3.3.2.1 Phénotypage des mutants

Deux mutations indépendantes ont été identifiées par TILLING pour le gène *MtLIL* dans les populations EMS de *M. truncatula* A17 présentes au laboratoire : les mutations P77S et E92K (Figure 3.7 et Tableau 3.2). L'analyse de ces mutants a été réalisée sur des plantes issues de deux rétrocroisements avec des plantes sauvages, les plantes homozygotes mutantes et sauvages ainsi que les hétérozygotes issus de ce deuxième rétrocroisement ont été utilisées pour l'analyse.

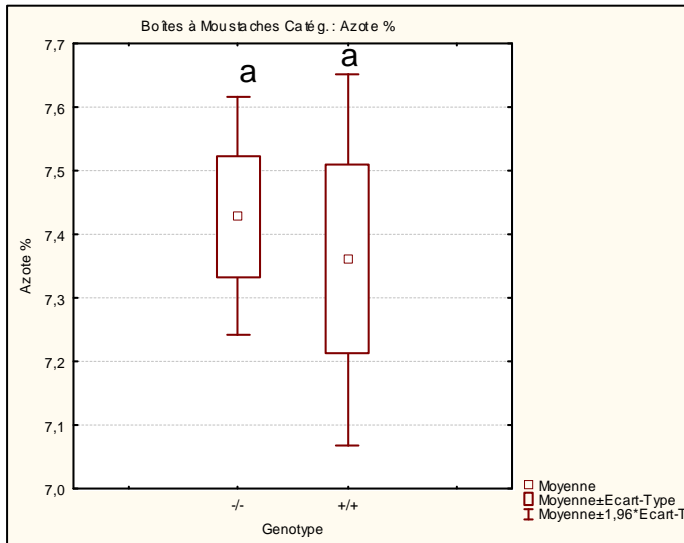
Comme précédemment avec les plantes transformées, le phénotypage des mutants obtenus a été réalisé, les mêmes caractères phénotypiques ont été suivis. En ce qui concerne les mutants P77S, aucune différence phénotypique n'a été mise en évidence entre les plantes mutantes homozygotes et sauvages (Figure 3.14) En revanche, les mutants E92K présentent une différence dans les poids de graine (Figure 3.15) : en effet, les

(a)

Génotype	Poids graine (mg)	N	Ecart-type
+/-	3,14	30	0,34
-/-	2,78	5	0,10
+/+	3,08	15	0,49

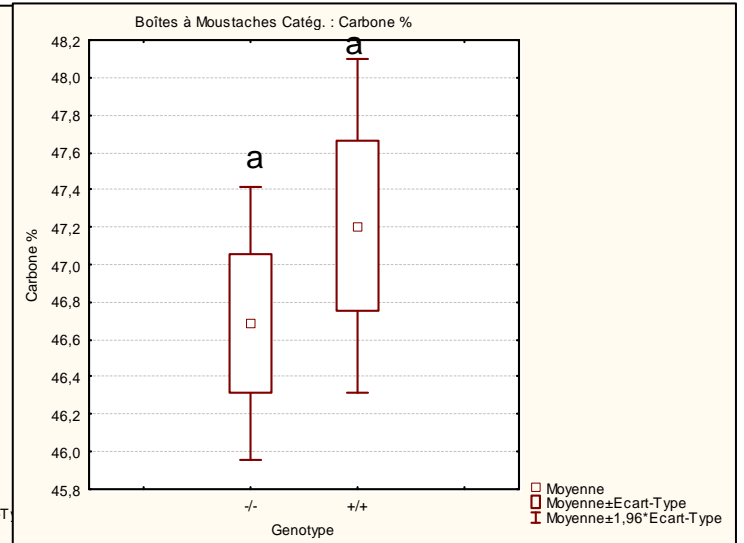


(b)



Génotype	Azote (%)	N	Ecart-type
-/-	7,42	3	0,09
+/+	7,36	6	0,14

(c)



Génotype	Carbone(%)	N	Ecart-type
-/-	46,68	3	0,37
+/+	47,20	6	0,45

**Figure 3.14 : Phénotypage des graines matures portant la mutation P77S**

(a) Poids de la graine mature

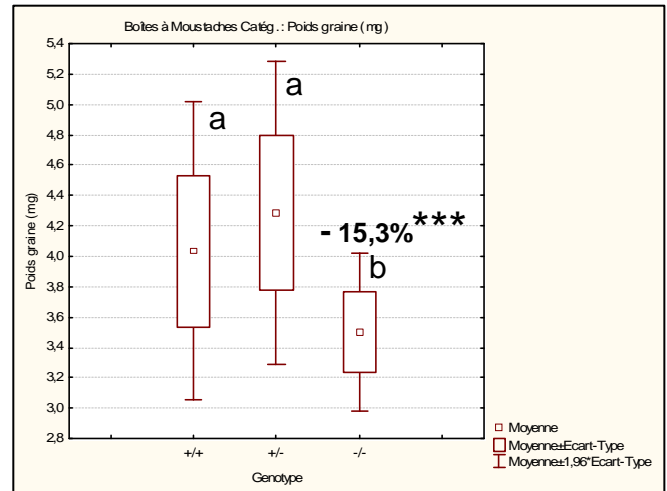
(b) Analyse du contenu en azote des graines

(c) Analyse du contenu en carbone des graines

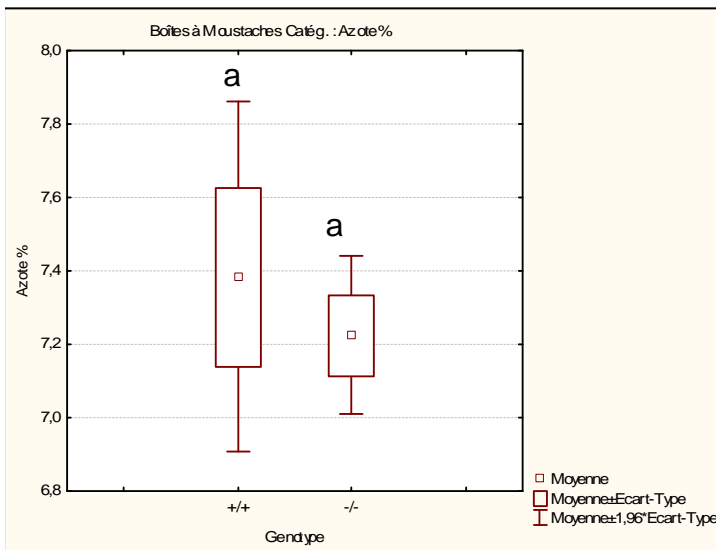
-/- homozygote mutant ; +/- hétérozygote et +/+ homozygote sauvage

(a)

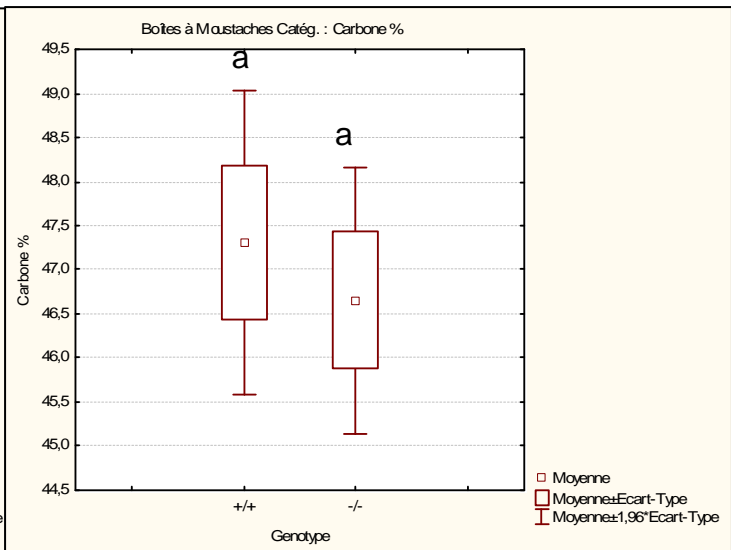
Génotype	Poids graine (mg)	N	Ecart-type
+/+	4,035667	20	0,499677
+/-	4,289467	15	0,507957
-/-	3,499000	10	0,265364



(b)



(c)



Génotype	Azote %	N	Ecart-type
+/+	7,383	12	0,243
-/-	7,224	6	0,109

Génotype	Carbone %	N	Ecart-type
+/+	47,309	12	0,879
-/-	46,657	6	0,770

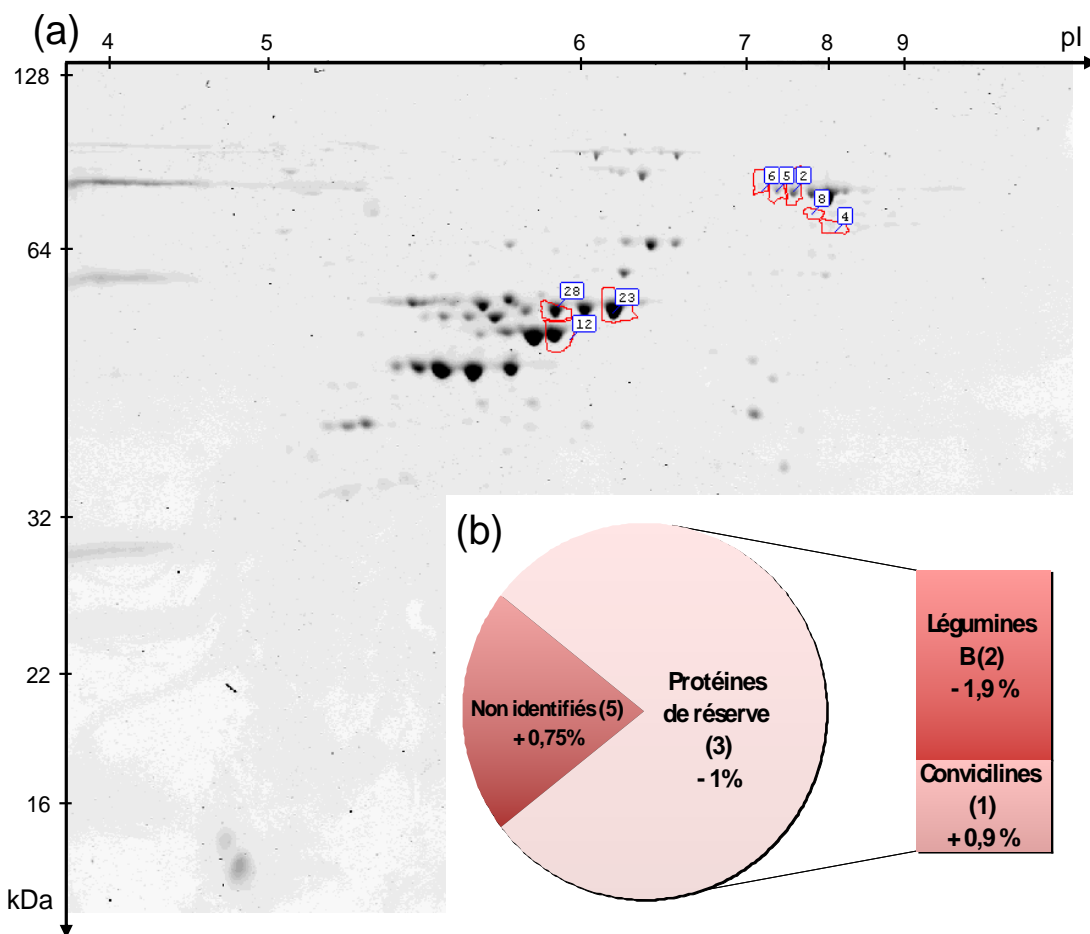
Figure 3.15 : Phénotypage des graines matures portant la mutation E92K

(a) Poids de la graine mature

(b) Analyse du contenu en azote des graines

(c) Analyse du contenu en carbone des graines

-/- homozygote mutant ; +/- hétérozygote et +/+ homozygote sauvage



**Figure 3.16 : Spots protéiques qui varient significativement dans les graines mutantes E92K**

- (a) Gel présentant les numéros de spots qui varient dans les graines mutantes et leur position  
 (b) Représentation des variations quantitatives par rapport au volume total du gel en fonction du type de protéines. Le nombre de spots concernés est indiqué entre parenthèses

N° Spot	Nom de la protéine	Numero d'accession	Classes fonctionnelles (MapMan)	Poids moléculaire et pI Théorique kDa / pI	Identification	Fold (par rapport au sauvage)	Abondance relative des spots ( en % du volume total)	
							WT	I1I
2	Convicilin	P13915	Development, Storage and LEA proteins	63,9 / 6,11	Gallardo et al. 2003 et 2007	3,0	0,46 ± 0,12	1,38 ± 1,52
4	Unknown					1,8	0,12 ± 0,05	0,21 ± 0,08
5	Convicilin	P13915	Development, Storage and LEA proteins	63,9 / 6,11	Gallardo et al. 2003 et 2007	1,8	0,3 ± 0,07	0,54 ± 0,14
6	Unknown					1,8	0,09 ± 0,02	0,16 ± 0,04
8	Unknown					1,5	0,06 ± 0,02	0,09 ± 0,02
28	Unknown					1,2	1,93 ± 0,18	2,25 ± 0,13
12	Legumin B	P16078	Development, Storage and LEA proteins	20,1 / 9,71	Gallardo et al. 2003 et 2007	0,7	3,54 ± 0,74	2,47 ± 0,47
23	Legumin B	P16078	Development, Storage and LEA proteins	20,1 / 9,71	Gallardo et al. 2003 et 2007	0,8	4,15 ± 0,68	3,32 ± 0,29

**Tableau 3.4 : Spots identifiés comme différentiellement exprimés dans les graines matures portant la mutation E92K dans la séquence *MtLIL*.**

Les numéros de spots se réfèrent à la figure 3.16

graines homozygotes mutantes présentent une diminution de leur poids d'environ 15% par rapport aux sauvages. En revanche, la mutation E92K n'affecte pas d'autres caractères de la graine tels que le contenu de la graine en carbone et azote (Figure 3.16).

### 3.3.2.2 Analyses biochimiques en 2D SDS PAGE

L'analyse protéomique des graines portant la mutation P77S n'a révélée aucune variation significative avec les graines sauvages ( $p$ -value < 5% et  $P > 75%$ ). Ce résultat confirme les résultats précédents de phénotypage où aucune différence n'avait également pu être identifiée. Il semble que cette mutation n'affecte pas la fonction du gène et donc ne modifie pas le développement de la graine.

En revanche, une analyse protéomique similaire sur les graines des plantes portant la mutation E92K a révélé une différence dans l'accumulation des protéines dans la graine mature. Huit spots protéiques varient significativement ( $p$ -value < 5% et  $P > 75%$ ) entre les graines mutantes et sauvages (Figure 3.16 et Tableau 3.4). La moitié de ces spots (4/8) ont pu être identifiés grâce à la carte de référence de la graine mature de *M. truncatula* A17 établie par Gallardo *et al.* (2003 et 2007). Tous les spots identifiés correspondent à des protéines de réserve : deux spots correspondent à des légumine B (spots #12 et #23) et deux spots à des convicilines (spot #2 et #5). Si on compare les abondances relatives de ces spots, on constate que les deux spots de légumine B sont moins abondants dans les graines mutées alors que les six autres spots, dont ceux de convicilines, sont au contraire plus abondants. De manière intéressante, on constate que la différence relative des spots sur et sous-accumulés est égale à zéro. Cette compensation entre les spots plus et moins abondants pourrait expliquer l'absence de variation du contenu en azote entre les graines mutantes et sauvages.

## 4. DISCUSSION

### 4.1 Identification de deux gènes de type LEC1 chez *M. truncatula*

Dans cette étude, nous avons caractérisé les deux premiers gènes codant pour des protéines de type HAP3 chez la légumineuse modèle *M. truncatula*. Les gènes HAP3 sont des sous-unités du CCAAT Binding Factor (CBF, Lothan *et al.* 1998), sous-unités très conservés durant l'évolution des eucaryotes et caractérisées par un domaine B. Chez *A. thaliana*, Kwong *et al.* (2003) ont mis en évidence l'existence de deux classes de HAP3 reconnaissables grâce à une signature différente d'acides aminés dans leur domaine B : les gènes de type LEC1 et ceux de type non-LEC1. L'analyse de séquence indique que les deux gènes identifiés chez *M. truncatula* (BI311277 et 1347.m00026) appartiennent à la même sous-classe, celle des gènes de type LEC1. Une analyse phénétique avec d'autres gènes codant des protéines de type HAP3 a confirmé cet assignement. De plus, cette analyse nous a permis de montrer que l'un de ces gènes est plus proche en séquence des gènes *LEC1*, nous l'avons donc appelé *MtLEC1* (1347.m00026), l'autre plus proche des gènes *LEC1-LIKE* (*LIL*), a été nommé *MtLIL* (BI311277).



L'analyse des profils d'expression révèle que ces deux gènes s'expriment spécifiquement dans la graine mais suivant deux profils différents. *MtLECI* s'exprime de manière plus précoce au cours de l'embryogenèse alors que *MtLIL* a un pic d'expression plus tardif. Cependant, la différence majeure entre ces deux gènes est l'intensité d'expression : en effet, on constate que *MtLIL* est beaucoup plus exprimé que *MtLECI*.

Ces constatations sur le léger décalage d'expression entre ces deux gènes sont comparables à ce qui a pu être décrit chez *A. thaliana* entre les gènes *AtLECI* et *AtLIL* (Kwong *et al.* 2003). En revanche, Kwong *et al.* (2003) ont montré que le gène *LIL* d'*A. thaliana* était exprimé dans la graine mais également dans les feuilles de rosette et dans les inflorescences. Dans notre étude, *MtLIL* est spécifique de la graine et ne présente aucune expression dans les organes végétatifs. En ce qui concerne la différence d'abondance des ARNm de ces deux gènes, ce résultat était également visible sur la figure 2 du manuscrit de Kwong *et al.* (2003), cependant il n'était pas mis en évidence, ni discuté dans ce travail. Cette différence est pourtant une donnée importante pour élucider le mécanisme d'action de ces deux gènes. Chez *M. truncatula*, au stade 12 DAP, correspondant au pic d'expression de *MtLECI*, on constate que *MtLIL* n'est pas à son maximum d'expression cependant les ARNm de *MtLIL* sont beaucoup plus abondants que ceux de *MtLECI*. On peut donc conclure que ces gènes sont exprimés en même temps au cours du développement de la graine, en revanche leur maximum d'expression présente un décalage de quelques jours.

## **4.2 Comparaison des résultats obtenus sur l'analyse des mutants**

Dans cette étude, nous avons identifié des plantes "perte de fonction" pour le gène *MtLIL* issues de deux procédés différents. Tout d'abord, nous avons transformé des plantes avec une construction de type RNAi afin d'inhiber l'expression du gène *MtLIL*. En parallèle, nous avons criblé une population de plantes EMS en utilisant la méthode TILLING. L'analyse des mutants, nous permet donc de comparer les résultats obtenus d'analyse fonctionnelle du gène *MtLIL* à partir de ces deux méthodes.

### **4.2.1 Extinction de gène par RNA interférence**

Le RNAi est une méthode puissante pour éteindre l'expression d'un gène mais peu utilisée chez *M. truncatula* en raison de la faible capacité de régénération *in vitro* de cette plante et particulièrement de son génotype de référence A17. Dans cette étude, nous avons utilisé le génotype R108 de *M. truncatula* dont l'une des caractéristiques est une meilleure régénération *in vitro*. Cette plante a d'ailleurs déjà été utilisée dans des approches de transgénèse chez *M. truncatula* pour générer une première population de mutants Tnt1 (d'Erfurth *et al.* 2003). Le premier problème auquel nous avons été confrontés a été le manque de données génomiques disponibles sur ce génotype, ce qui a fortement ralenti nos investigations. De la même manière, l'identification des spots sur gels 2D a également été retardée en raison de la difficulté à utiliser la carte de référence du





protéome de la graine de A17 pour identifier les spots chez R108. Cette différence nous a conduit à séquencer les spots différentiellement accumulés entre les graines transformées et sauvages.

Cette méthode nous a quand même permis d'obtenir deux plantes transformées III contenant la construction en "épingle à cheveu" et présentant une baisse d'expression du gène *MtLIL*. Ces plantes ont été analysées et présentaient peu de différences phénotypiques au niveau des graines alors que nous pensions qu'au contraire le phénotype serait de type léthal sauf pour quelques allèles plus faibles comme décrit par Kwong *et al.* (2003) chez *A. thaliana*. Cette différence observée avec l'étude de Kwong où les embryons étaient fortement affectés jusqu'à avortés dès les stades précoces de l'embryogenèse peut s'expliquer principalement de deux manières :

Premièrement, nous ne pouvons pas exclure qu'une autre copie du gène existe chez *M. truncatula*, comme c'est le cas chez *A. thaliana*, ou alors qu'un autre gène fonctionnellement redondant soit capable de partiellement compléter l'inhibition d'expression du gène *MtLIL*.

Deuxièmement, une explication serait de dire que l'inhibition de l'expression de *MtLIL* n'est pas suffisante pour affecter le développement de la graine. En effet, les niveaux d'extinction du gène *MtLIL* observés dans ces transformants n'ont pas dépassé les 75% (baisse de 74% pour L73 et de 50% pour L31) alors que ce type de construction est décrit dans la littérature pour entraîner une baisse d'expression du gène supérieure à 90% de l'expression (Wesley *et al.* 2001)). Cette hypothèse est vraisemblable car dans l'étude de Kwong *et al.* (2003) où les auteurs utilisaient également des transformants de type RNAi, les transformants possédant des niveaux intermédiaires d'ARNm de *AtLIL* présentaient un phénotype "sauvage" et les plantes transformées affectées dans le développement embryonnaire ne montraient pas un niveau détectable de transcrits *LIL*. En conséquence, on peut penser que dans notre étude, les transformants ne présentaient pas une baisse suffisante d'ARNm de *MtLIL*. Pour expliquer ce constat deux problèmes d'ordre techniques sont à prendre en considération et viennent directement du choix du promoteur que nous avons utilisé dans cette expérience. En effet, le choix du promoteur constitutif et fort 35S du virus de la mosaïque du chou-fleur s'est imposé de lui-même pour cette construction car couramment utilisé chez les plantes. Cependant, de récentes publications ont montré que premièrement, ce promoteur, dit constitutif, ne le serait pas durant les premières étapes de l'embryogenèse chez *A. thaliana* (Kwong *et al.* 2003). Deuxièmement, une comparaison entre des promoteurs forts et constitutifs à montrer que chez *M. sativa*, une variété de luzerne très proche de *M. truncatula*, le promoteur 35S ne présentait pas une très forte activité en comparaison d'autres promoteurs dits "forts" (Samac *et al.* 2004).

#### 4.2.2 Identification de mutants pour *MtLIL* par criblage TILLING

Le criblage TILLING de la population EMS a permis de mettre en évidence deux mutations différentes dans le gène *MtLIL*, P77S et E92K. Les deux mutations obtenues correspondent à des modifications ponctuelles de type SNP (substitution de G en A) susceptibles de modifier la conformation de la protéine et donc d'affecter sa fonction. Contrairement aux plantes transformées RNAi, les mutants ont été identifiés dans une population de mutants du génotype de référence A17, ce qui constitue un avantage pour leur analyse. En raison du fort effet



mutagène du traitement qui génère plus de 1000 mutations dans leur génome des plantes traitées. Nous avons pris soin de réaliser deux rétrocroisements avec des plantes sauvages afin d'enlever les mutations "accessoires" présentes dans le génome. Aussi, nous disposons de plusieurs plantes homozygotes sauvages et mutantes provenant de rétrocroisements différents afin d'obtenir des plantes contenant un "bruit de fond" de mutations différent. Cette hétérogénéité du fond génétique des différentes plantes nous permet de fiabiliser nos résultats, en revanche cela génère une importante variabilité intra-génotypes nécessitant de grands effectifs pour les analyses statistiques. Les analyses phénotypiques et biochimiques suggèrent que seulement l'une des deux mutations affecte la fonction de la protéine. Ce résultat ne nous permet donc pas de disposer d'une série allélique de mutants souhaitable dans une analyse de mutants EMS. L'analyse fonctionnelle parallèle de plusieurs mutants "perte de fonction" aurait renforcée les résultats obtenus.

#### **4.3 Rôle de *MtLIL* lors du développement de la graine**

Après identification des deux gènes de type LEC1 chez *M. truncatula*, nous avons focalisé sur le gène *MtLIL* afin de valider son rôle lors du développement de la graine. L'analyse du profil d'expression de ce gène en rapport avec les profils d'expression des gènes codant les principales protéines de réserve chez *M. truncatula* montre une corrélation entre le maximum d'expression du gène et le début de synthèse des vicilines et convicilines. Cette corrélation peut laisser supposer une relation fonctionnelle entre l'expression du gène *MtLIL* et l'expression des gènes codant les globulines 7S. Cette hypothèse est renforcée par trois résultats obtenus dans ce travail. Tout d'abord, le gène *MtLIL* présente une forte expression dans l'embryon, lieu de synthèse des globulines. Ensuite, dans les graines transformées l11, on observe une légère augmentation du contenu en azote (de 2%, p-value = 0,08). Cette augmentation est corrélée à une légère augmentation de la concentration en protéines totales dans ces mêmes graines transformées. Enfin, on constate une modification du protéome de la graine mature chez les graines de la lignée RNAi *MtLIL*, avec une augmentation significative de plusieurs spots protéiques correspondant à des globulines. Cette modification du contenu protéique affecte principalement les convicilines si l'on raisonne en nombre de spots qui varient, et en particulier des formes de convicilines ayant subi des modifications post-traductionnelles ; en revanche si l'on raisonne en quantité relative de spots, ce sont les légumines et les vicilines qui semblent les plus affectées. Ce dernier point est en partie confirmé dans les graines des plantes mutées E92K, où l'on constate également que la mutation du gène *MtLIL* affecte l'accumulation de convicilines, vicilines et légumines. De plus, dans ces mutants on observe que la mutation E92K entraîne la baisse du poids des graines matures chez les mutants.

En conclusion, ces deux approches "perte de fonction" ont révélé des phénotypes communs, cependant il est encore plus informatif de comparer les phénotypes qui diffèrent. Ces deux méthodes affectent le gène *MtLIL* de manières différentes : le RNAi induit une baisse d'expression du gène sans altération de la structure protéique alors que la mutation ponctuelle des mutants EMS affecte la conformation de la protéine mais pas forcément son niveau d'expression. La baisse d'expression du gène induit une légère hausse de l'accumulation de protéines de réserve dans la graine mature tandis que la mutation du domaine de fixation des sous-unités entraîne une baisse



du poids de la graine mature. En conclusion, on peut émettre l'hypothèse que le gène *MtLIL* est un gène impliqué dans un processus précoce du développement de la graine qui régule indirectement le remplissage et donc le contenu protéique des graines ainsi que leur poids à maturité.

#### **4.4 Hypothèses sur le rôle des gènes de type LEC1**

L'analyse des séquences protéiques entre les deux gènes *HAP3*, *MtLIL* et *MtLEC1*, indique une forte homologie du domaine B (Tableau 3.1 et Figure 3.3) déjà décrite chez *A. thaliana* (Kwong *et al.* 2003). En effet, des expériences de complémentation de mutants *lec1* grâce à des protéines chimériques ont été réalisées et ont permis de montrer le rôle essentiel du domaine B pour la fonction des gènes de type HAP3 (Lee *et al.* 2003). Ce rôle du domaine B dans la fonction de LEC1 a été confirmé par des expériences de complémentation de mutants *lec1* à l'aide de la séquence codante de *LIL* sous le contrôle des séquences bordantes de *LEC1* (Kwong *et al.* 2003 ; Yazawa *et al.* 2004 et Alemanno *et al.* 2007). Ces observations indiquent que ces deux protéines ont la même fonction et sont donc capables de fixer les mêmes séquences cibles d'ADN ainsi que les mêmes sous-unités co-activatrices. En revanche, le gène endogène *LIL* ne peut pas substituer fonctionnellement *LEC1* dans les mutants *lec1*, ce qui implique un rôle fondamental des séquences régulatrices promotrices et 5',3' non traduites dans la fonction de ces deux gènes.

Les hypothèses expliquant que le gène endogène *LIL* ne peut pas compléter des mutants *lec1* sont (i) une expression temporelle différentes entre ces deux gènes, (ii) une expression spatiale différente, (iii) une intensité d'expression de *LIL* insuffisante et enfin (iv) une régulation croisée entre les gènes *LIL* et *LEC1* par l'intermédiaire des régions régulatrices (Kwong *et al.* 2003).

Dans notre étude, nous montrons que *MtLEC1* s'exprime en même temps que *MtLIL* même si un léger décalage des maximums d'expression est observé (Figures 3.5) ce qui exclut l'hypothèse d'une expression temporelle différente. De plus, nous constatons une expression de *MtLIL* plus forte que celle de *MtLEC1* (Figure 3.6), ce qui exclut également l'hypothèse d'une intensité insuffisante de *MtLIL*. En revanche, nous ne pouvons pas conclure sur l'hypothèse d'une expression spatiale différente, en effet même si ces deux gènes sont exprimés dans les mêmes tissus, l'embryon et l'albumen, nous n'avons pas réussi à déterminer leur expression cellulaire (Figure 3.8). Enfin, la dernière hypothèse d'une régulation entre ces deux gènes est appuyée par l'observation que nous avons faite sur la dérégulation de l'expression du gène *MtLEC1* dans les transformants sous-exprimant *MtLIL*.

Même si nos résultats ne nous permettent pas d'exclure toutes les autres hypothèses, une régulation négative du gène *LIL* par le gène *LEC1* expliquerait la plupart de nos résultats. En effet, *LEC1* est connue pour jouer un rôle de régulateur positif dans la synthèse des protéines de réserve par action indirecte via *FUS3* et *ABI3* (Kagaya *et al.* 2005a). Dans notre étude nous constatons à l'inverse qu'une diminution d'expression de *MtLIL* entraîne une



augmentation significative de certains spots correspondant à des protéines de réserve des graines, ce qui indiquerait un effet négatif de ce gène sur la régulation des protéines de réserve. L'une des hypothèses envisageables sur le rôle des gènes de type LEC1 serait de dire que *LEC1* et *LIL* ont des rôles antagonistes. Ainsi, en suivant notre raisonnement, une balance *LEC1/LIL* pourrait réguler la transition entre l'embryogenèse et le remplissage de la graine.





## 5. CONCLUSION DU CHAPITRE

L'analyse du rôle des facteurs de transcription dans le développement des plantes est un domaine en plein essor. Le développement de la graine n'échappe pas à ce constat et de nombreuses publications apparaissent sur ce sujet. De nombreux TFs identifiés comme clefs lors du développement de la graine chez *A. thaliana* sont à l'étude chez cette plante modèle, cependant la composition et la régulation de ses protéines de réserve diffèrent par rapport aux légumineuses. Les gènes de type LEC1, identifiés et décrits chez *A. thaliana*, semblent jouer des rôles clefs à toutes les étapes du développement de la graine. Dans cette étude, nous avons (1) identifié des gènes de type LEC1, chez la légumineuse modèle *M. truncatula*, (2) validé fonctionnellement le rôle de *MtLIL* et enfin (3) proposé une hypothèse sur le rôle de ces gènes dans le développement de la graine et plus particulièrement pendant la phase de remplissage.

### Synthèse des résultats

Cette étude a permis dans un premier temps d'identifier deux gènes de type LEC1, *MtLECI* et *MtLIL*, chez la légumineuse *M. truncatula*. Ces deux gènes qui s'expriment au début de la synthèse des protéines de réserve semblent impliqués dans la régulation de ces dernières. En effet, il a été montré que le gène *LECI*, chez *A. thaliana* est un régulateur de la synthèse des protéines de réserve via la régulation des gènes *FUS3* et *ABI3* (Kagaya *et al.* 2005a). Quant au gène *MtLIL*, nous avons montré son implication dans l'accumulation des protéines pendant le remplissage et dans le déterminisme du poids de la graine mature à l'aide de mutants EMS et/ou transformants RNAi. En s'appuyant sur le travail de Kwong *et al.* (2004), notre étude permet d'émettre l'hypothèse d'une régulation antagoniste des gènes *LECI* et *LIL* lors de la transition embryogenèse-remplissage. Dans ce modèle hypothétique de régulation, *LECI* aurait un rôle positif sur l'activation de la synthèse des protéines de réserve tandis que *LIL* à l'inverse aurait un rôle négatif.

### Perspectives de ce chapitre

Le travail présenté dans cette partie ne tient pas compte de l'analyse des mutants *MtLECI* et utilise donc les données de la littérature disponibles chez *A. thaliana*. Les perspectives de ce travail seront tout d'abord l'analyse fonctionnelle des mutants EMS pour le gène *MtLECI* identifiés dans cette étude. De plus, une analyse du double mutant *lec1-III* est également envisageable. De ces deux analyses dépendra la validité de notre modèle de régulation des gènes de type LEC1 chez *M. truncatula*.



**CHAPITRE 4 : Rôle du gène *PsL1L*  
dans l'élaboration de la qualité de  
la graine de pois (*P. sativum*)**



## CHAPITRE 4

### Rôle du gène *PsLIL* dans l'élaboration de la qualité de la graine de pois (*P. sativum*)

#### 1. OBJECTIFS ET DEMARCHE

Chez le pois, peu de gènes directement impliqués dans la régulation de la qualité finale du grain ont pu être identifiés. En effet, même le "gène de Mendel", qui permit il y a plus d'un siècle de mettre en évidence les lois régissant l'hérédité des caractères ouvrant ainsi la voie à la génétique, n'a été identifié que récemment (Bhattacharyya *et al.* 1990). De plus, de nombreuses données sont aujourd'hui accumulées sur les différentes espèces modèles mais ne seront peut-être jamais validées, ni exploitées sur les espèces cibles correspondantes.

Dans le chapitre précédent, nous avons identifié deux gènes de type LEC1 chez la légumineuse modèle *M. truncatula*. Ces deux gènes sont exprimés pendant le développement de la graine au début de la synthèse des protéines de réserve. Nous avons d'ailleurs montré que l'un des deux gènes, *MtLIL*, jouait un rôle dans le remplissage du grain. En effet, les graines transformées présentaient une teneur en azote légèrement plus importante et surtout une augmentation du volume de plusieurs spots protéiques par rapport aux graines sauvages. Sur la base de ces résultats, nous avons décidé d'étendre notre analyse à une légumineuse d'intérêt agronomique, *P. sativum*, afin de tester l'influence de l'orthologue du gène de *M. truncatula*, le gène *PsLIL*, sur le développement de la graine de pois et surtout son impact sur la qualité finale de la graine.

Nous avons donc voulu confirmer les observations faites chez *M. truncatula* tout en élargissant les connaissances sur la régulation transcriptionnelle du développement de la graine chez le pois. L'objectif de ce travail a donc été :

- (i) d'identifier le gène *PsLIL* chez le pois malgré le peu de ressources génomiques disponibles,
- (ii) d'identifier des mutants du gène *PsLIL* afin de réaliser une caractérisation fonctionnelle et d'élucider son rôle durant le développement de la graine et dans l'élaboration de la qualité finale du grain,
- (iii) d'étudier la variabilité naturelle de la séquence de ce gène afin de déterminer si le polymorphisme de séquence de ce gène est associé à certains caractères liés à la variabilité de la qualité de la graine de pois.

Pour y parvenir, nous disposons de deux populations de plantes de pois : tout d'abord la population de plantes mutées à l'EMS générée au laboratoire (INRA UMRLEG Dijon) et dont le criblage avec la méthode TILLING nous a permis de mettre en évidence des plantes présentant des mutations dans le gène *PsLIL* (Collaboration



avec INRA URGV, Evry). Ensuite, la deuxième population essentielle à cette analyse a été la collection de pois gérée par le Centre de Ressources Biologiques (CRB, INRA UMR-LEG Dijon). Cette collection de 372 écotypes très contrastés va nous permettre d'étudier le polymorphisme de séquence du gène *PsL1L* grâce à la méthode d'EcoTILLING (Comai *et al.* 2004).

Ce chapitre est constitué d'une première partie sur l'identification de *PsL1L* et sur sa caractérisation fonctionnelle chez *P. sativum* en utilisant des mutants identifiés par criblage TILLING. Ces résultats seront présentés sous la forme d'une publication en cours de soumission. Puis, dans une seconde partie, nous discuterons des relations que nous avons identifiées entre le polymorphisme de séquence de *PsL1L* et des caractères de qualité du grain en utilisant la technique d'EcoTILLING.





## **2. IDENTIFICATION ET CARACTERISATION FONCTIONNELLE DU GENE *LEAFY COTYLEDON1-LIKE (PsLIL)* CHEZ LE POIS.**

### ***LEAFY COTYLEDON1-LIKE*, a regulator of seed storage protein synthesis in pea**

#### Résumé :

Les protéines de type LEC1 sont connues chez *A. thaliana* pour jouer un rôle clef dans les différents processus du développement de la graine dont l'embryogenèse, le remplissage et la dessiccation. Dans cette étude, nous avons identifié un homologue du gène *LEAFY COTYLEDON1-LIKE* chez le pois, que nous avons appelé *PsLIL*, en raison d'une séquence protéique très conservée avec la protéine correspondante chez *A. thaliana*. Ce gène est exprimé au début du remplissage et apparaît être indirectement impliqué dans la synthèse des protéines de réserve des graines. Nous avons montré que les lignées mutantes pour le gène *PsLIL*, identifiées par un criblage TILLING, présentaient une plus grande quantité de protéines de réserve que les sauvages. Cette augmentation de l'accumulation des protéines de réserve n'affecte pas de spots protéiques spécifiques mais se traduit plutôt par une augmentation globale du contenu protéique et en particulièrement celui des légumineuses. Nous proposons donc un rôle de régulation négative du gène *PsLIL* sur la synthèse des protéines de réserve, rôle antagoniste à celui du gène *PsLEC1*.



## **LEAFY COTYLEDON1-LIKE, a regulator of seed storage protein synthesis in pea**

Jérôme Verdier, Christine Le Signor, Françoise Moussy, Myriam Sanchez, Delphine Héricher, Abdel Bendahmane, Karine Gallardo and Richard Thompson

### **ABSTRACT**

LEC1-type proteins are known in *A. thaliana* to play a key role during several processes of seed development including embryogenesis, seed filling and desiccation tolerance. In this study, we identified a pea homologue of *LIL*, that we named *PsLIL*, exhibiting a highly conservative amino acid sequence with the corresponding *A. thaliana* protein. This gene is expressed at the onset of seed filling and appears to be indirectly involved in seed storage protein synthesis. We showed that lines bearing two different mutations in the *PsLIL* gene displayed higher seed storage protein contents than do wild-type seeds. This increase of storage protein accumulation does not affect specific spots but rather the global accumulation of proteins and especially those corresponding to the legumin class. We propose a negative regulatory role for *PsLIL* in storage protein synthesis and an antagonistic effect on the *PsLEC1* gene.



## INTRODUCTION

Legumes played important roles in the development of agriculture and, the associated emergence of civilization. They are distinguished by their ability to form nitrogen-fixing symbioses with soil bacteria and thus to grow in the absence of mineral or organic nitrogen supply (Graham and Vance, 2003). All geneticists know the key role that a legume, the garden pea (*Pisum sativum*) played in the history of genetics thanks to Mendel's experiments. However, fewer will appreciate the reasons why Mendel grew pea in his garden. In Europe, pea represents the most cultivated pulse crop (UNIP-Prolea, <http://www.prolea.com/>) and mainly serves as a protein source for human and animal nutrition. Seeds of grain legumes are nutritionally valuable due to their high content of proteins, fibres and oils (Wang, 2003). Its seed protein content, being less than that of other pulses such as soybean, is the subject of breeding efforts. Whereas pea seed content represents an important economic interest, mechanisms implied in storage compound synthesis are largely unknown and few genes involved in determining mature seed quality have been identified (Bhattacharyya *et al.* 1990).

Accumulation of seed storage compounds occurs essentially in cotyledons during the seed filling phase, following embryogenesis (Meinke *et al.* 1981; Walling *et al.* 1986; Ladin *et al.* 1987; Pang *et al.* 1988). The final phase of seed maturation is characterized by desiccation that permits the embryo to enter into metabolic quiescence. During the seed filling phase, pea seeds accumulate storage macromolecules that ultimately serve as a nutrient source for the germinating seedling. These storage macromolecules represent the most important economic and nutritional components. Pea seeds are mainly composed of starch (approximately 50% of seed dry weight) and protein (approximately 25% of seed dry weight). The protein fraction is essentially composed of storage proteins that are classified into two families: the albumins (also named 2S albumins according to protein size range) and the multigene globulin family, composed of vicilins, convicilins (also named 7S globulins) and legumins (12S globulins), which are accumulated sequentially during seed development (Gatehouse *et al.* 1982 and Boulter *et al.* 1987).

Seed developmental transitions, including the phase of accumulation of seed storage protein are mainly regulated at transcriptional level in the legume model, *Medicago truncatula* (Gallardo *et al.* 2007 and Verdier *et al.* 2008). Moreover, several studies in *A. thaliana* and cereals have shown the importance of transcription factors during seed development: regulators of maternal contribution to seed coat (*TTG2*, Garcia *et al.* 2005), regulators of embryo morphology (*MONOPTEROS*, Hardtke et Berleth, 1998), factors governing the deposition of seed storage proteins (*OPAQUE2*, Unger *et al.* 1993) and lipids (*WRINKLED1*, Cernac and Benning, 2004), and those directing other aspects of seed maturation (e.g. desiccation tolerance, *ABSCISIC ACID-INSENSITIVE5*, Finkelstein and Lynch, 2000). Other transcription factors, such as *LEAFY COTYLEDON1 (LEC1)*, *LEAFY COTYLEDON2 (LEC2)*, *FUSCA3 (FUS3)* and *ABSCISIC ACID-INSENSITIVE3 (ABI3)* play key roles during embryogenesis, seed filling and desiccation (Lotan *et al.* 1998; Harada *et al.* 2001; Yazawa *et al.* 2004 and Kagaya *et al.* 2005ab). Embryo development in the *lec1*, *lec2*, *fus3* and *abi3* mutants is abnormal in both morphogenesis and maturation processes. These mutants exhibit defects in desiccation tolerance and in synthesis/accumulation of storage materials. For example, the *LEC1* gene is sufficient to specify cotyledon



identity during embryogenesis and to confer embryonic characteristics to seedlings (Lotan *et al.* 1998). Other studies showed that ectopic expression of LEC1 is sufficient to induce somatic embryo formation from vegetative cells, suggesting a key position for this gene in the regulatory hierarchy controlling embryogenesis (Fambrini *et al.* 2007). Moreover, LEC1 plays a key role during seed filling since expression of the storage protein genes is greatly reduced in *A. thaliana* mutant plants (Meinke *et al.* 1994; West *et al.* 1994 and Parcy *et al.* 1997). A recent study (Kagaya *et al.* 2005a) suggested that LEC1 controls expression of the seed storage protein genes in a hierarchical manner, involving ABI3 and FUS3. Given the central role of this gene in the control of seed development, we asked if the LEC1 gene also encodes a regulator of seed quality in a crop plant. LEC1 shares similarity with the Heme-Activated Protein3 (HAP3) subunit of CCAAT binding factor (CBF). CBFs are eukaryotic transcription factors, also known as NF-Y in yeasts or vertebrates (Edwards *et al.* 1998), constituted by a hetero-oligomeric complex of at least three subunits: HAP2, HAP3 and HAP5 (HAP4 is the fourth subunit only present in yeast). HAP3 subunits are characterised by their B domain, a region constituted of a subunit binding site and a DNA binding site that is conserved across eukaryotic organisms. Multiple HAP3 genes exist in *A. thaliana* and are divided into two different classes: the LEC1-types and the non LEC1-type genes. These two classes differ by 16 conserved amino acids present in the B domain (Kwong *et al.* 2003). Out of 16 amino acid residues, one, an aspartate residue, is sufficient to confer partial LEC1 activity to a non LEC1-type B domain. (Lee *et al.* 2003). In *A. thaliana*, the HAP3 subunits most closely related to LEC1 belong to LEC1-type. Different LEC1-type genes have been identified in *A. thaliana* or other species and are divided into LEC1 gene and LEC1-LIKE (L1L) genes (Kwong *et al.* 2003; Yazawa *et al.* 2004; Fambrini *et al.* 2007 and Alemanno *et al.* 2008). LEC1 and L1L genes have distinct functions in embryogenesis but each can substitute functionally one another, suggesting that their B domain have a basic function in embryogenesis.

Here, we identified and isolated the first HAP3 subunit LEC1-type from the agronomic plant *Pisum sativum* that we named *P. sativum* LEAFY COTYLEDON1-LIKE (*PsL1L*) according to sequence homology. The *PsL1L* expression profile was determined, which suggested a role during the onset of seed storage protein synthesis. *PsL1L* mutants were identified from an EMS population using TILLING screening. Phenotyping and biochemical analysis of mutant seeds allowed us to predict a role for this gene in storage protein synthesis. Finally, a hypothesis about the possible role of L1L and LEC1 in filling seeds will be discussed.

## MATERIALS AND METHODS

### Plant Material

*Pisum sativum* plants (cultivar 'Cameor') were grown in two-litre pots filled with sterile pouzzolane in greenhouse. Plants were automatically watered with a nutrient solution of 3.5N/3.1P/8.6K. Temperature was above 14°C during the night and below 30°C during the day, complementary artificial lighting was provided to obtain a 16 hour day. The plants were not inoculated with *Sinorhizobium sp.* bacteria and nitrogen supply was designed to be mainly mineral and non-limiting. Individual flowers were tagged on the day of their opening





corresponding to one day after pollination. For each stage of the three biological replicates, developing seeds were collected on ice to prevent any dehydration, from pods harvested from 10 to 36 Days After Pollination (DAP). Flowers and leaves were collected at flowering. Before DNA and RNA extraction, all seed and plant tissue samples were frozen quickly and separately ground in liquid nitrogen using mortar and pestle. Each powder was homogenized and stored at -80°C prior use.

#### Identification of *PsLIL* genomic sequences

We used the amino sequence of *A. thaliana* *LEAFY COTYLEDON1-LIKE* gene (AtL1L, Kwong *et al.* 2003) as a query to identify related *M. truncatula* polypeptides. Genomic researches were carried out using the basic local alignment search tool (BLAST) program on MtGI v.8 database. Primers were designed on sequence of *M. truncatula* *LIL* putative gene in order to amplify a related pea sequence from developing seed cDNA. PCR fragments obtained were cloned and sequenced (MWG Sequencing service, Germany) leading us to identify a fragment of *PsLIL* cDNA. Several rounds of 5' RACE-PCR were performed to obtain the full length cDNA using the rapid amplification of cDNA ends system (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions. RACE-PCR amplification products were cloned using the TOPO® Cloning Kit (Invitrogen, Carlsbad, USA) and sequenced (MWG Sequencing service, Germany) to obtain the genomic fragment containing two exons.

#### RNA isolation and qRT-PCR analysis

Total RNA was isolated from developing seeds at 10, 12, 14, 16, 20, 30, 36 DAP according to the method described by Chang *et al.* (1993). RNA concentration was measured in a Gene Quant Pro spectrophotometer and RNA integrity was checked on a 1.5% (w/v) RNase-free agarose gel. Approximately 4µg of total RNA was treated with RNase free DNase RQ I™ (Promega, Madison, USA) for 45 min according to the manufacturer's instructions. Genomic DNA specific primers were used as controls to confirm the absence of genomic DNA in RNA extracts. RT reactions were performed using iScript one-step reverse-transcriptase kit (Bio-Rad, Hercules, USA) according to the manufacturer's instructions. Other cDNA preparations from mature seeds and from other plant tissues were synthesized as described previously (add ref.). All samples were measured in technical duplicates on three biological replicates harvested and extracted independently. Master mixes were prepared: each samples contained Mesa Green supermix (Eurogentec, Seraing, Belgium), gene-specific primers (200 nM) and 20 ng cDNA. *HISTONE H1* (#X05636) and *ELONGATION FACTOR BETA* (EF1β, #X96555) were previously tested to be constitutive throughout pea seed development (data not shown). The geometrical mean of these both internal standards were used to normalize expression of different genes. Controls lacking cDNA were included and the specificity of primers designed for qRT-PCR was confirmed by melting curve analysis. Real-time PCR reactions were performed on a Roche LC480 (Roche Applied Science) using standard conditions: 95°C for 2 min, followed by 40 cycles of incubation at 95°C (30 s), 60°C (1 min). Melting curves were recorded

Name	Forward	Tm	Reverse	Tm
PsL1L	GCGTAAGATTCTACCTCCACATGC	60	CCTTTGCTCTCTTTGACATCGTTC	60
seed maturation prot PM34	GAGGAGCACTGGTTGGAAGC	60	CGATGAAGAGAGCTGGGCA	60
P54 protein	GAACTCACAACATGCAAAGACCA	60	CCATGCATATCCGTTTGTCTTC	60
Albumin 2 PA2	TTGTGAGAAGCAAACGCTGC	60	TCGCAACGGATTTCTTGT	60
Legumin A2	CGCATTGAAGGGACGTGCAAG	60	GTGACTTTGCAGCCACAGCATAG	60
Vicilin 47kD protein.	TGTTCTCTTCCTTGTCAATTTCTT	60	CCGAAGGAAAAGGAGATTTGA	60
Legumin B Fragment).	TCTTGCAAATGCTTTCGGACT	60	GGCCACGGTTCCATTAGACT	60
Convicilin precursor.	GGTATCACCGCGCTCAAGAT	60	TGTAGTCCTCAATGGGAAAGCC	60
FUS3	GGAGACTTTGTTAACATGCATCG	60	ATCATTGGTTGTGTGAGTTCTCTC	60
histone H1	GCCAAGCCAAAACGTGTTGC	60	CGGAGTTGTCTTCACAGATGTCT	60
polyubiquitine	GCCAAGATTCAAGACAAGGAAGG	60	ACCACGAAGTCGCAACACAA	60

**Table S1:** List of genes profiled in qRT-PCR with annotations, primer sequences and experimental data.

from 55°C to 95°C, reading every 0.5°C. For quantification of gene expression, data were analyzed using the PCR standard curve method with the LC480 gene expression software. All gene accessions, primer pair sequences and qPCR conditions are available in the Supplementary Table S1.

#### Sequence alignment analysis

Identification of 26 other HAP3 genes was carried out using previous studies (Kwong *et al.* 2003; Yasawa *et al.* 2004; Fambrini *et al.* 2006; Alemanno *et al.* 2008) and researches in database using BLAST algorithm. The amino acid sequences were aligned using clustalW (Thompson *et al.* 1994), then a phylogenetic analysis was performed using CLC Workbench v.4 software (CLC Bio software). The neighbour joining method of clustering was used., A bootstrap analysis with 500 replicates was performed to test robustness.

#### Isolation of *PsL1L* mutants from TILLING screening and mutant phenotyping

*PsL1L* mutants were identified from an EMS population containing 4800 *Pisum sativum* lines using Target-Induced Local Lesions In Genome (TILLING) screening (McCallum *et al.* 2000). Plant growth conditions, genomic DNA extraction, PCR and digestion were performed as described in Dalmais *et al.* (2008). To confirm mutations, PCR amplification products were sequenced (MWG Sequencing service, Germany) and sequence analysis (Chromas v.1.45 software) was performed.

The pea stock centre for tilling mutants (INRA-UMRLEG) provided us with a batch of 4 M3 seeds for each requested mutant. All M3 progenies were back-crossed twice with the reference line Cameor to remove most of the influence of background mutations. Selfed progenies from the back-cross (BC2) provide plants segregating for the mutant allele suitable for mutation-phenotype co-segregation analysis. A comparison of wild-type to mutant plants using different back-crossed progenies best controls for background mutations while phenotyping. For each mutant, finding a homozygous plant in a heterozygous M3 progeny requires a genotyping step. DNA from each of the M3 plantlets was isolated. For the SNP detection, we used routinely the derived Caps strategy (<http://helix.wustl.edu/dcaps/dcaps.html>). Homozygous plants were phenotyped during plant growth in the greenhouse. To determine seed dry weight and water content during development, three pools of ten fresh seeds were weighed just after harvest and after drying at 80°C for 48 h for each genotype. To evaluate total carbon and nitrogen content of seeds, we performed Dumas analysis (Allen, 1974) using a CHN analyser (Carbo-Elba).

#### FPLC analysis of seed protein composition

Proteins were extracted from 100 mg of seed powder (20 min. shaking at 450 rpm; centrifugation 30 min., 6000 g, 4°C; drying of the pellet overnight under aspiration). Then, pellets were solubilized in 2.5 ml extraction buffer

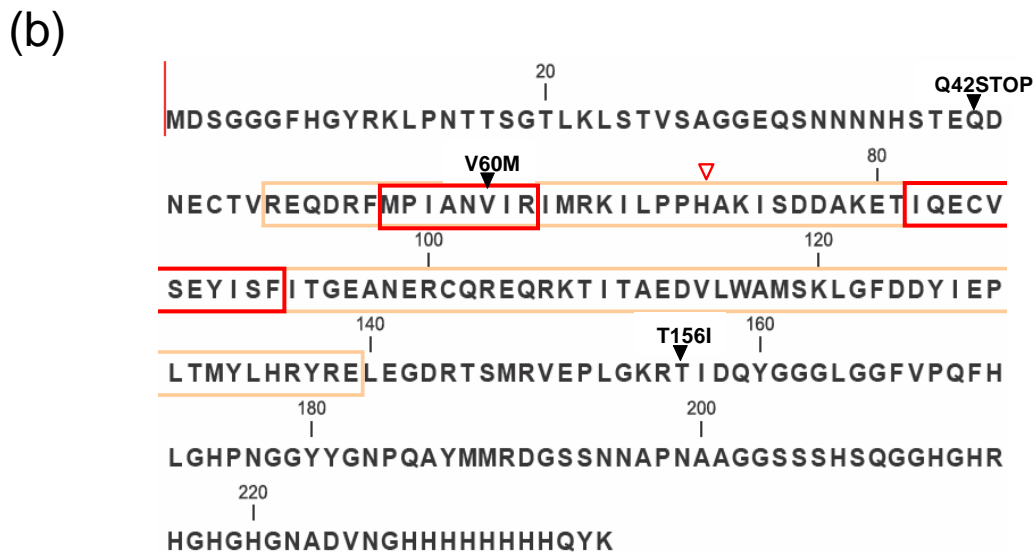
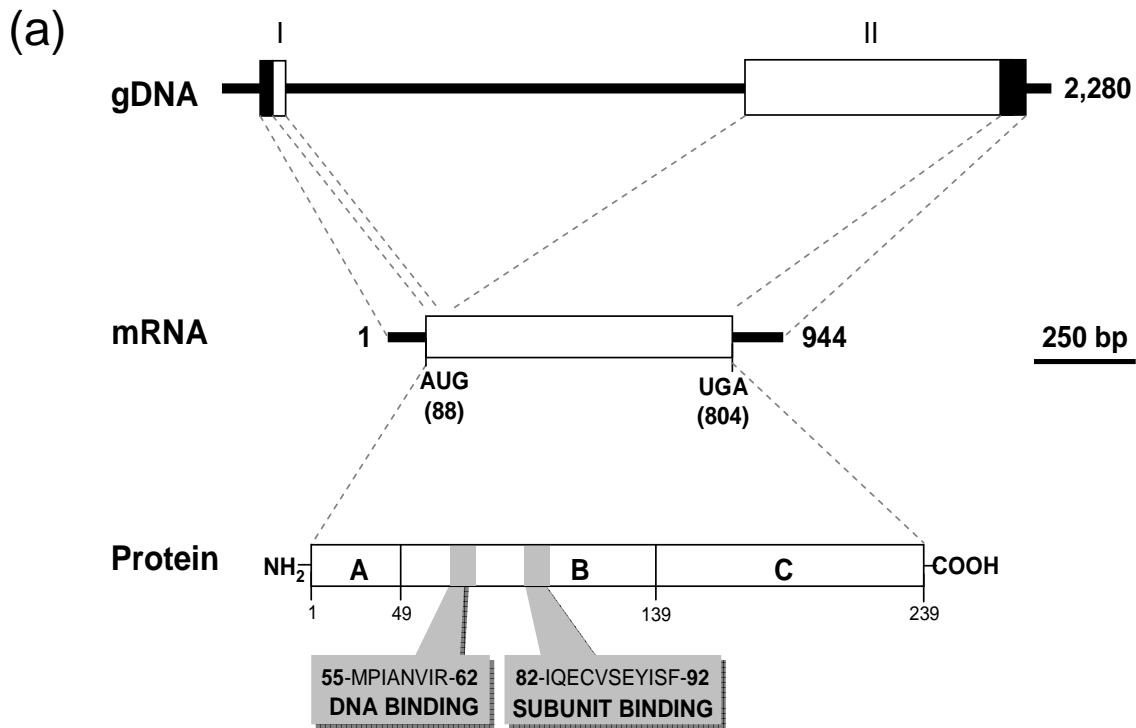


(Tris 50 mM, K<sub>2</sub>SO<sub>4</sub> 5%, EDTA 1 mM, β-mercaptoethanol 0.05%, benzamidine 1 μM, leupeptine 1 μM, PMSF 1 μM, pH 8.5) for one hour on a shaker then centrifuged (6000 g, 20 min., 4°C). Samples were then filtered on 0.22 μm cellulose acetate syringe-filters (Millipore Corp., Milford, MA). Fast Protein Liquid Chromatography was performed with a Waters (Waters Europe S.A.S., Saint Quentin en Yvelines, France) system (refs. 600E system controller, 717 injector, 486 detector) according to Baniel *et al.* (1998), with volumes of 75 μl of filtered protein extract. We used a MonoQ 5/50 GL analytical column (Amersham Biosciences). The NaCl gradient for protein elution was 0-0.72 M (40 min., 1 ml/min). Protein detection was performed by measuring UV absorbance at 214 nm. The quantification of each FPLC fraction was performed by integrating peak areas.

Collected fractions corresponding to vicilin/convicilin were purified and desalted through centrifugal concentrators equipped with 5000 MWCO (Molecular Weight Cut Off) polyethersulfone membranes according to the manufacturer's instructions (Vivaspin6, Vivascience AG, Hannover, Germany). Quantities of 7 and 14 μg of protein purified samples from collected fractions were separated on a 12% acrylamide gel. The quantification of 1-DE gel bands was performed by densitometry analysis using Odyssey software (LI-COR, Nebraska, USA).

#### 2-DE analysis of seed protein composition

Duplicate of all protein extractions were carried out. Total mature seed proteins were extracted from 20 mg of powder in 2 ml of lysis buffer (7.5 M urea; 1.1% Pharmalyte pH 3 to 10; 63 mM CHAPS (Amersham Biosciences, Saclay, France); 22 mM trizma hydrochloride; 17.3 mM trizma base (Fluka); 2.2 M thiourea; 0.25% (v/v) triton X-100 (Sigma-Aldrich); protease inhibitor cocktail complete mini1X (Roche Diagnostics). After shaking (1000 rpm, 30 min., 4°C), 8 mM DTT, 60 U DNaseI, 5.8 Kunitz unit/mg RNaseA were added, and another shaking step was performed (1000 rpm, 20 min., room temperature). Samples were then centrifuged (20000 g for 10 min., 4°C) and supernatant was collected and centrifuged one more time (20000 g for 20 min., 4°C). Protein concentrations in the protein extracts were measured according to Bradford (1976) by using a reagent (BioRad Protein Assay; Bio-Rad) compatible with the reducing and denaturing agents used to prepare the protein extracts. Two-dimensional separation of seed proteins was performed as described in Gallardo *et al.* (2002) with minor modifications. Isoelectrofocusing of 150 μg of proteins was performed in an IPGphor apparatus (Amersham Biosciences) using 24-cm gel strips (non linear pH gradient from 3 to 10, Amersham Biosciences) in a total volume of 450 μl of rehydration buffer (7 M urea, 2 M thiourea, 65 mM CHAPS, 20 mM DTT, Pharmalyte pH 3 to 10, 0.5%, 2% Triton X-100). After isoelectrofocusing, the strips were equilibrated, and SDS-PAGE separation was done using the Isodalt System (Amersham Biosciences). One 2-DE gel per biological replicate was performed. Gels were stained with Coomassie Brilliant Blue G250 (Biorad) according to the procedure described by Mathesius *et al.* (2001), with minor modifications: gels were stained for three days in the staining solution and destained for 24 h. After several washing steps in MilliQ water (Millipore Corp., Milford, MA), gels were scanned with an Odyssey Infrared Imager electronic scanner (LI-COR, Nebraska, USA) and then stored in a 20% [w/v] ammonium sulphate solution. After gel digitization, images were imported into Progenesis SameSpot software (Non Linear Dynamics, Newcastle, UK) to be analysed. To compare differences



**Figure 1: Sequence analysis of *PsLIL***

(a) Diagrammatic representations of the structure of the *P. sativum LIL* gene, the spliced mRNA and the deduced translation product. Localization of exon-intron junctions was indicated. Exons are represented by white boxes with numbered regions; intron is represented by lines and 5' and 3' non-transcribed regions with black boxes. In mRNA representation, the white box represents regions of exons I and II and black lines correspond to the untranslated sequences. Within the protein, different domains of *PsLIL* are indicated with their positions in amino acid sequence. The conserved B domain is represented including DNA binding and subunit binding with corresponding consensus motif.

(b) Deduced amino acid sequence of the *PsLIL* gene. The B domain is boxed and DNA binding and subunit binding are indicated with a red background. Positions of different mutations identified in this gene are indicated with arrows and corresponding mutant names.

in protein abundance among the different genotypes, a one-way ANOVA and a Student-Newman-Keuls test were performed for each spot. Spots were defined as differentially expressed according to a p-value less than 5% and a test power more than 70%.

### Protein spot identification

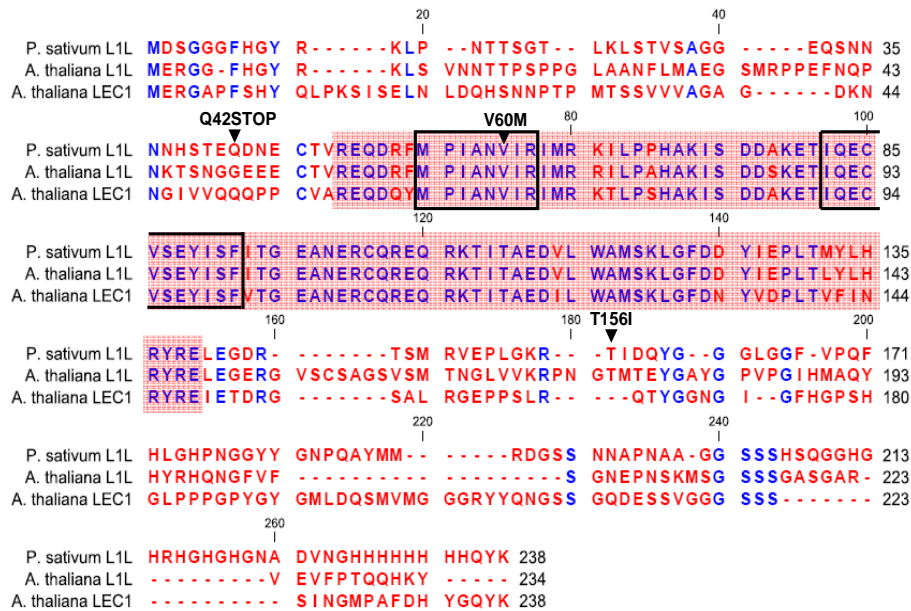
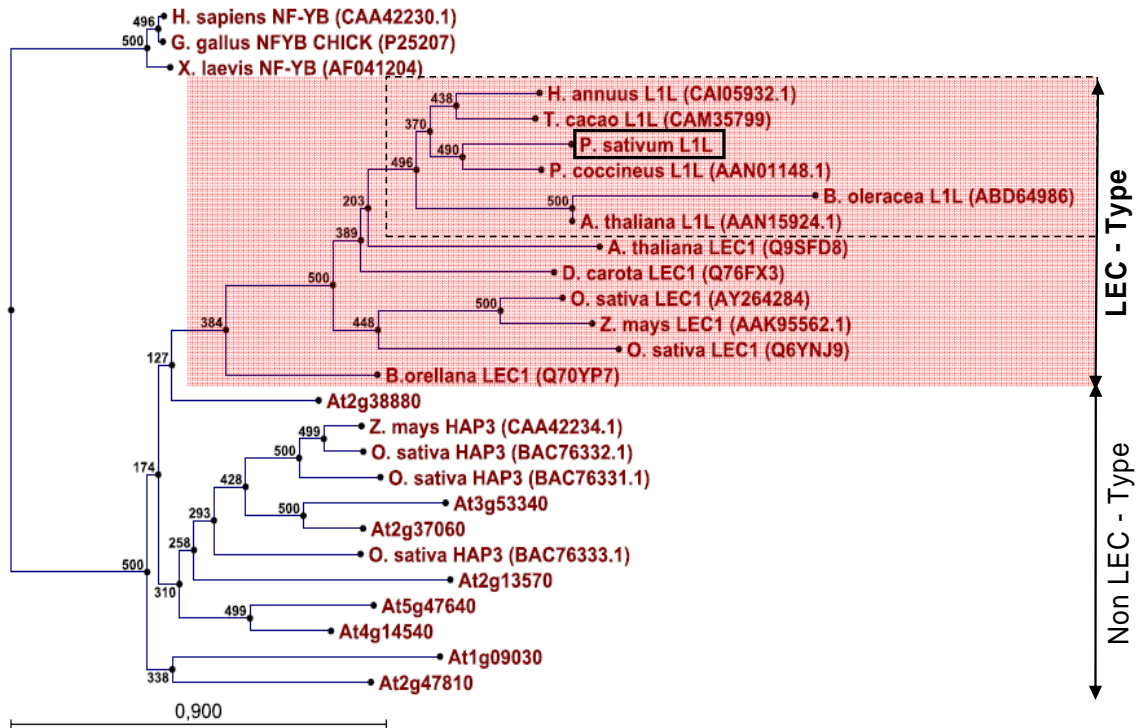
To identify spots from 2D gels, we referred to the pea seed map established by Bourgeois *et al.* (submitted) (available at <http://www.inra.fr/legumbase/peaseedmap>). We also identified new spots by mass spectrometry using an ion trap nanoLCMS/MS at INRA Nouzilly, France (nanoHPLC Ettan MDLC system, GE Healthcare, Germany, and LTQ, Thermo Electron, USA). Public databases (TIGR Medicago v.8 and NCBI) and a private database included pea seed cDNA (Künne *et al.* 2005) were used for protein identification.

## RESULTS

### Identification and sequence analysis of the *PsL1L* gene

Based on the putative *M. truncatula LIL* gene (unpublished results), we amplified a fragment of pea orthologue from pea seed cDNA. The 428 bp PCR fragment obtained possesses a sequence closely related to the 3' terminal sequences of *M. truncatula LIL* (*MtLIL*) cDNA. Using 5' RACE-PCR, we obtained the full length cDNA sequence containing 944 nucleotides and respectively the 87 bp and 140 bp of the 5' and 3' untranslated regions. The predicted open reading frame, a coding sequence of 717 bp, encodes 239 amino acids. The predicted molecular weight of the polypeptide is 26.5 kDa, and the protein has a theoretical isoelectric point of 6.69 ([http://www.expasy.ch/tools/pi\\_tool.html](http://www.expasy.ch/tools/pi_tool.html)). This polypeptide has an amino acid sequence closely related to AtL1L, (54 % of identity) and to AtLEC1 (46 % of identity). It also shows 95% sequence similarity in the B domain, between amino acids 49 to 139, which contains the DNA binding site (55 MPIANVIR 62), the subunit binding site (82 IQECVSEYISF 92) and the consensus sequence specific to the LEC1-type subclass of the HAP3 proteins (Figure 1a and 1b). All 16 typical amino acids characteristic of the LEC1-type subclass are conserved, including an essential aspartic acid residue at position 84, which confers LEC1 activity to the B domain (Kwong *et al.* 2003; Lee *et al.* 2003). Close examination of the central B domain sequence alignments showed 95 % of protein sequence identity with the AtL1L B domain (85% with AtLEC1 B domain) and 100% of identity to the *MtL1L* B domain (unpublished results). We named this sequence *P. sativum LEC1-LIKE* (*PsL1L*). To analyse the correspondence of *PsL1L* with HAP3 proteins from other species, a phylogenetic analysis comparing protein sequences from a range of eukaryotes was performed. The eukaryotic HAP3 protein classification can be divided firstly into two groups containing HAP3 proteins from animals and plants; secondly in plants, we can divided





**Figure 2: Analysis of HAP3 subunits in various species**

(a) Neighbor joining tree was generated using CLC sequence viewer v.4. The tree topology was obtained with a bootstrap analysis of 500 replicates. In brackets are indicated the GeneBank accession numbers. LEC-type proteins were highlighted using a red background. L1L proteins were separated of LEC1 type with a dotted box. PsL1L is indicated with a black box.

(b) Amino acid sequence alignment of PsL1L was compared with AtLEC1 (Q9SFD8), AtL1L (AAN15224). Residues highlighted in blue and red represent respectively identical and non identical amino acids. The central B domain is highlighted in red background including DNA binding and subunit binding regions, which are indicated in black boxes. Arrows indicate positions of different mutations.

two subgroups: the LEC type and the non-LEC type (Figure 2). The plant LEC type and non-LEC type represent two well-separated subgroups, except for one protein (At2g38880). Within the plant LEC type group, the L1L subclass and the LEC1 subclass are clearly distinguishable and *PsLIL* is more closely related to the L1L class than to LEC1.

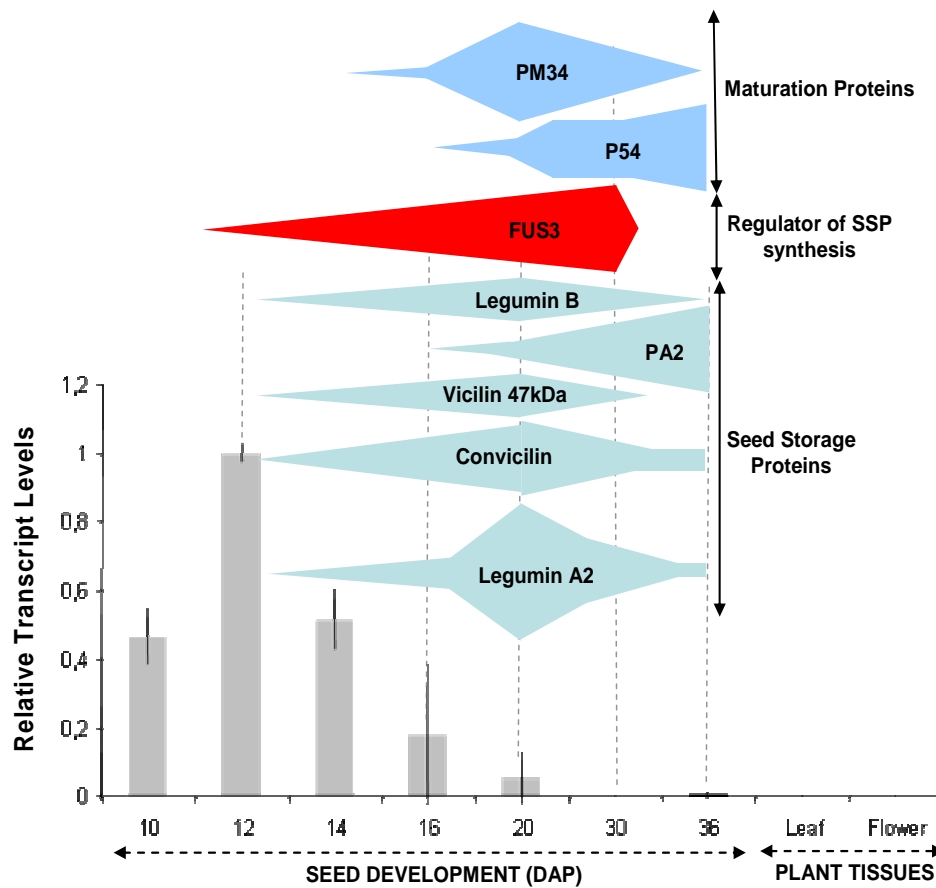
Finally, a comparison between the *PsLIL* cDNA and the genomic DNA amplified using a *PsLIL* specific primer set revealed that the coding region of the *PsLIL* gene is interrupted by an intron of 1336 bp (Figure 1a).

#### *PsLIL* is highly expressed during seed development

The *PsLIL* gene was identified within a pea seed cDNA database implying expression in seed. A qRT-PCR amplification confirmed the seed expression of *PsLIL* mRNA suggesting a role during seed development (Figure 3), only background levels being detected elsewhere, in flowers and leaves. To better define the role of *PsLIL* during seed development, transcript profiling of different seed storage proteins (vicilin, convicilin, legumin B and albumin), a regulator of storage protein synthesis (FUS3, Kagaya *et al.* 2005b) and maturation proteins (PM34 and P54) were performed in developing seed using qRT-PCR. *PsLIL* transcript was detected at 10 Days After Pollination (DAP) reaching a maximum at 12 DAP, and it decreased gradually until no signal was detectable at 20 DAP (Figure 3). Its peak of expression is thus related to the onset of seed storage protein expression. These results suggest that *PsLIL* has a role at the onset of seed filling.

#### Identification and phenotyping of *PsLIL* mutants

To further characterize the role of *PsLIL* in seed development, we looked for *PsLIL* gene mutants in an Ethyl Methane Sulfonate (EMS) mutant population containing 4800 pea plants (cultivar 'Cameor') (Dalmais *et al.* 2008). Twenty independent point mutants were identified which correspond to two silent mutations and 18 non-silent or missense mutations. Of the 18 missense mutations, nine mutants encode amino acid substitutions that affect protein function based on Sorting Intolerant From Tolerant software (SIFT) predictions (<http://blocks.fhrc.org/sift/SIFT.html>, Ng *et al.* 2003) and one exhibits a premature stop codon located at the beginning of the conserved B domain. We have focused on three mutants: the stop codon mutant (i.e. substitution of a glutamine by a stop codon, Q42STOP) corresponding to a truncated protein version lacking the B and C domains, and two missense mutants (Figure 1b). The first mutant (V60M) displayed mutation in the DNA binding site of the highly conserved B domain. The second mutant (mutant T156I) displayed mutation located in the C domain (Figure 1b). These two mutations affect amino acids that may have deleterious effects on gene function: V60M mutation corresponds to a substitution by a similar amino acid group but is located in a



**Figure 3: Expression profiles of *PsLIL* gene during developing seeds.**

qRT-PCR amplification of *PsLIL* during developing seeds and in plant tissues. RNA corresponding to different seed storage proteins, maturation proteins and regulators of seed storage protein synthesis were profiled in seeds using qRT-PCR. All gene accessions and primer sequences are available in supplementary data Table S1.

	Measured Parameter	Stop Mutants			V60M Mutants			T156I Mutants								
		WT		I1I	WT		I1I	WT		I1I						
Plant	Plant length (cm)	68.2 ± 18.1	NS	73.0 ± 15.9	NOT MEASURED											
	Number of seed per pod	3.51 ± 0.66	NS	3.79 ± 0.81												
Developm- ental Seed	20 DAP	Dry weight per seed (mg)	119.05 ± 38.86	NS							114.99 ± 25.59					
		Water content (%)	66.65 ± 2.12	NS							64.79 ± 2.89					
	30 DAP	Dry weight per seed (mg)	163.84 ± 24.30*	-18.5%							133.46 ± 37.61*					
		Water content (%)	14.06 ± 10.92	NS							11.71 ± 6.60					
Mature Seed	Fresh weight per seed (mg)	219.6 ± 11.0**	-15.6%	184.5 ± 18.2**							215.7 ± 4.1	NS	199.3 ± 1.3	218.1 ± 6.5	NS	214.3 ± 2.5
	Dry weight per seed (mg)	183.2 ± 9.6**	-15.9%	154.1 ± 15.7**							177.8 ± 4	NS	165.2 ± 1.1	179.0 ± 6.1	NS	176.1 ± 2.1
	Water content (%)	16.6 ± 0.2	NS	16.5 ± 0.3							17.6 ± 0.3*	-2.9%	17.1 ± 0.1*	17.9 ± 0.3	NS	17.8 ± 0.1
	Nitrogen content (%)	4.09 ± 0.03***	+9.5%	4.48 ± 0.1***							4.17 ± 0.19	NS	4.25 ± 0.06	3.99 ± 0.05***	+14.8%	4.58 ± 0.02***
	Carbon content (%)	44.24 ± 0.56	NS	44.19 ± 0.6	42.74 ± 2.02	NS	43.97 ± 0.52	44.07 ± 0.76	NS	44.18 ± 0.58						
	Germination test (% seed viability)	93.3	NS	100	100	NS	100	100	NS	100						

Mean values are significantly different from the corresponding wild-type at the 5% level are annotated with \*; 1% with \*\* and 0.001 with \*\*\* (Student's t test)  
NS means Non Significantly different

**Table 1: Phenotyping of different pea *I1I* mutants**

Traits with mean values with significant differences between wild-type and mutant plants are indicated in red according to a t-test analysis.

highly conserved amino acid sequence. The T156I mutation corresponds to a change in polarity and hydrophobicity of amino acid, possibly affecting protein conformation.

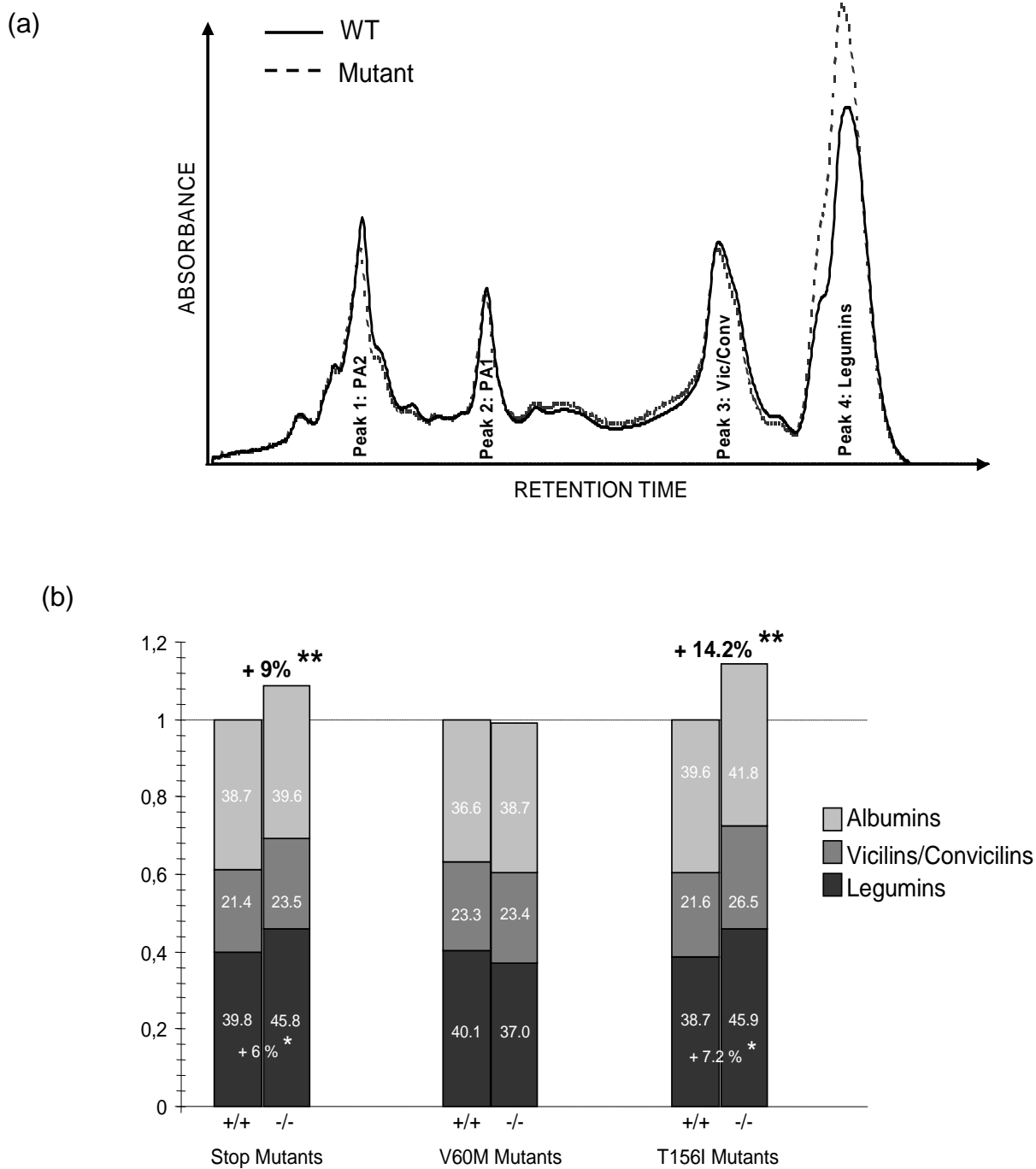
Homozygous mutant plants from the three different alleles exhibited a normal reproductive and vegetative development (Table 1). We analyzed seed fresh weight, dry weight, and water content during seed development in different mutant lines. Measurement of dry and fresh weight revealed differences only between the stop codon mutant and corresponding wild type. Stop codon mutant plants exhibited 15% smaller mature seeds than wild-type plants. Water content was not affected reflecting that seed developmental kinetic is not affected.

Using the Dumas method, carbon to nitrogen ratio was measured in mature seeds and revealed significant differences between mutant and wild-type seeds. Interestingly, mutant seeds have approximately 9.5% higher nitrogen content when the protein is truncated and 14.8% higher in T156I mutants compared to wild-types. A slight decrease in carbon content observed in mutant seeds compared to wild-types was not statistically significant (Table1).

#### Analysis of seed protein content and composition in mature seeds

To examine the basis for the increase in nitrogen content, seed protein content and composition were analysed by Fast Protein Liquid Chromatography (FPLC). The chromatograms obtained from protein extracts of pea mature seeds exhibited four main peaks (peak #1 to peak #4, Figure 4). As previously described (Baniel *et al.* 1998), the peaks correspond to seed storage protein families with respectively albumins in peaks #1 and #2 (respectively PA2 and PA1), convicilins/vicilins in peak #3 and legumins in peak #4 (Figure 4). The FPLC analysis confirmed that mutants exhibit changes in mature seed protein content. Except for the V60M mutant, which displayed similar protein content to the wild-type, the other seed mutants both had significantly higher protein content higher than wild-type (p-value < 0.005) with respectively 9% and 14% higher total protein abundance for the stop and T156I alleles. This result is consistent with the increase of nitrogen content observed in mutant seeds using the Dumas method (Table 1).

The highly contrasted profiles, indicating quality in the separation of the protein families, allowed quantitative analyses of each storage protein family. According to peak area quantification, V60M mutants did not show any change in protein composition of mature seeds. This result confirms that this mutation does not affect Lec1-like protein function. In the other mutants, seed protein composition was affected, in particular legumin abundance. A t-test showed a significant increase of the legumin class in mutant seeds (p-value < 5%) with respectively 6% and 7.2% increase in the stop and T156I mutants, although this increase is not sufficient to explain the total increase in seed protein content. Even if differences were observed in vicilin and convicilin content, they were not statistically significant. The vicilin/convicilin class were examined in more detail by taking the protein complements of the corresponding peak (#3) and fractionating on 1-DE gels. Quantification of 1-DE gel bands did not reveal differences in accumulation of these two classes of storage proteins between wild-type and mutant seeds (Data not shown).



**Figure 4: FPLC analysis of mature seeds of *III* mutants and their corresponding wild-types**

(a) Chromatogram of pea mature seeds with different peaks corresponding to seed storage proteins. This chromatogram is an example of the result obtained with a STOP codon mutant to illustrate the highly contrasted profiles of the different protein families.

(b) Quantification of protein families of different FPLC fractions. Quantifications were performed by integrating different peak areas. Significantly different mean values between mutants and wild-types are indicated at the 5% level (annotated with \*) and at the 1% level (annotated with \*\*).

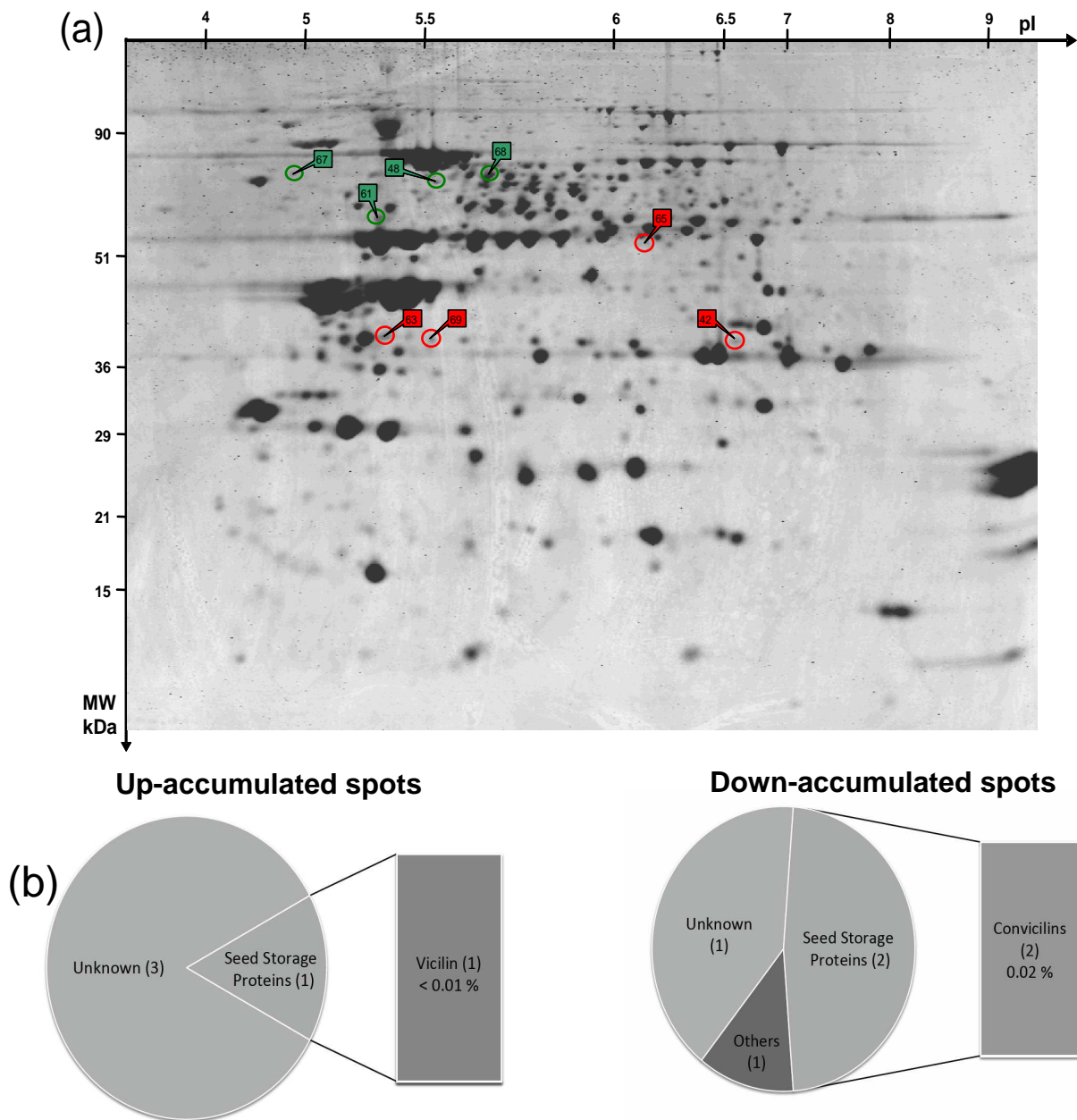
### Comparison of mutant and wild-type proteomes of mature pea seeds

Given that *lil* mutants accumulate higher seed protein content, especially higher legumin content than wild-types, we carried out a comparison of the mature pea seed protein complement using 2D-SDS PAGE between two independent *lil* stop mutants and their corresponding wild-types. This analysis distinguishes a global increase of protein spot abundance from an increase in specific spots or protein classes in mutants. Out of approximately 400 spots detected in our 2-DE analysis, only eight spots showed significant variations (p-value < 0.05 and power > 70%) in abundance between the wild-types and stop codon mutants (Supplementary Figure S1). According to peptide identification, four of these spots (4/8) were identified (Supplementary Table S2). We have analysed in detail the protein abundance changes between *lil* mutants and respective wild types. Four spots accumulate less in mutants than in wild-type seeds, including one vicilin (spot #69), and four spots were more abundant in mutant seeds including two convicilins (spots #48 and #68) and a protein disulfide isomerase (spot #67). The total value of down and up-accumulated spots represent less than 0.01% of the total spot volume in mature seed. These minor variations did not explain the modification of the storage protein accumulation in mutant seeds observed using FPLC and Dumas analyses, implying a global regulation of storage protein spots rather than effects on specific polypeptides.

## DISCUSSION

### Identification of a HAP3 subunit of the CCAAT binding factor (CBF)

In this study, a pea homolog of the *HAP3* gene of *A. thaliana*, *AtL1L*, was identified and isolated. We have named *PsL1L*. In *A. thaliana*, HAP3 subunits are divided into two distinct classes: LEC1 and non-LEC1 types (Kwong *et al.* 2003), which differ in the amino acid sequence of the B domain. *PsL1L* shares the 16 amino acid residues which confer LEC1-type signature (Gusmaroli *et al.* 2001) including the essential aspartic acid residue in position 84 (corresponding to Asp-55 in *AtLEC1*). In *A. thaliana*, Asp-55 is critical to confer partial B domain activity of the CBF complex (Lee *et al.* 2003). *PsL1L* gene encodes a protein with sequence similarity to HAP3 subunit of the CCAAT binding domain (CBF, also known as NF-Y) (Lotan *et al.* 1998). The HAP3 subunit consists of three regions: the A, B and C domains, with the central B domain conserved throughout eukaryotic evolution (Lotan *et al.* 1998). The B domain possesses amino acids required for interaction with DNA and with other CBF complexes (Kim *et al.* 1996 and Sinha *et al.* 1996). These results suggest that *PsL1L* is a member of a eukaryotic CBF complex and shares homology to the HAP3 LEC1-types. A phylogenetic analysis showed that *PsL1L* is more closely related to *AtL1L* than to *AtLEC1* (Figure 2). qRT-PCR amplification experiments indicate that *PsL1L* mRNA was highly detected in developing seeds similarly to *AtL1L* transcripts (Kwong *et al.* 2003). High levels of *PsL1L* mRNA were detected at the transition between embryogenesis and seed filling. This expression pattern is closely related to the onset of vicilin and legumin synthesis (Boulter *et al.* 1982) and could be correlated to the hypothesis that LEC1-type HAP3 subunits control the expression of seed storage protein genes in a hierarchical manner (Kagaya *et al.* 2005a).



**Figure S1: 2-DE analysis of pea *lil* mutants.**

(a) 2-DE gel of pea mature seed with spots up-accumulated in STOP mutants compared to wild-types in red and down-accumulated in green. Spot accession numbers are indicated to refer to Table S2.

(b) Pie chart representing relative abundance with respect to percentage of total volume of up and down-accumulated spots. The number of spots present in each class is indicated in brackets.

Spot number	Molecular Weight et experimental pI kDa / pI	Protein name	Accession	Functional Classes (MapMan)	Molecular Weight et theoretical pI kDa / pI	MS Tools	Fold (par rapport au sauvage)	Abundance relative (% total volume)	
								WT	l1l
42	37751 / 6.4	Unknown					1,5	0,022 ± 0,005	0,032±0,002
63	38278 / 5,3	Unknown					1,2	0,066 ± 0,004	0,081± 0,009
65	54884 / 6.19	Unknown					1,2	0,034 ± 0,002	0,041± 0,004
69	37750 / 5.5	Vicilin Precursor	P13918	Development, Storage and LEA proteins	52231 / 5.4	nanoLC-ion trap	1,1	0,049 ± 0,002	0,055±0,003
68	78211 / 5.7	Convicilin	Q9M3X6	Development, Storage and LEA proteins	68926 / 5.42	nanoLC-ion trap	0,9	0,081 ± 0,003	0,071± 0,003
61	63968 / 5.3	Unknown					0,8	0,087 ± 0,008	0,07± 0,003
67	78211 / 4.9	Protein disulfide-isomerase precursor (PDI)	P29828	Protein folding	57089 / 5	nanoLC-ion trap	0,8	0,033 ± 0,002	0,028± 0,002
48	75513 / 5.5	Convicilin	Q9M3X6	Development, Storage and LEA proteins	68926 / 5.42	nanoLC-ion trap	0,7	0,034 ± 0,001	0,024± 0,005

**Table S2: List of proteins differentially accumulated between mutants and corresponding wild-types and their relative abundance.**



### Role in seed development

The role of *LEAFY COTYLEDON1-LIKE* genes has been investigated in *A. thaliana* (Kwong *et al.* 2003), in *H. annuus* (Fambrini *et al.* 2006) and in *T. cacao* (Alemanno *et al.* 2008). In these studies, authors focused on the role of these genes in zygotic or somatic embryogenesis. In this study, we isolated a homologue of *LIL* in pea and investigated its influence on seed development and composition using pea mutants identified with TILLING screening in an EMS population. As in *A. thaliana*, the analysis of the mutant phenotypes revealed that vegetative development was not affected by the *lil* mutation. However, unlike in *A. thaliana* where phenotypes of RNAi mutants non expressing *LIL* genes were observed in reproductive development and especially during seed development (Kwong *et al.* 2003), *lil* mutants in *P. sativum* were viable including those from the null mutation (i.e. homozygous stop mutants). Moreover, desiccation/germination processes of pea seed mutants were not affected. This difference in phenotypes suggests that *PsLIL* is likely to be substitutable by other genes in pea. This hypothesis is supported by the existence of at least two different *LIL*-class genes in the *A. thaliana* genomic database.

### Focus on the implication of *PsLIL* in mature seed composition

LEC genes are well known to regulate seed development including the filling phase (Meinke *et al.* 1994; West *et al.* 1994; Parcy *et al.* 1997; Kwong *et al.* 2003). To study the role of *PsLIL* genes in the regulation of seed storage protein accumulation, we analysed pea mature seeds in wild-type and in *lil* mutants. Mutant seeds have higher nitrogen contents than wild-type seeds. FPLC analysis confirmed that this increase in nitrogen content reflects directly a higher protein content in mutant seeds which is largely due to storage proteins. Moreover, this approach reveals major differences in accumulation of globulin types (11S and 7S). Legumin content augments but this does not account for the total increase in storage proteins, whereas albumin and vicilin/convicilin contents are not statistically different (Figure 4). In conclusion, *PsLIL* appears to be an inessential gene during embryogenesis, which has the effect of lowering protein content through its action on storage protein accumulation. The *lil* mutation is therefore potentially of interest for improving pea seed quality by conferring higher protein content on mature seeds.

To analyse in more detail the effect of *PsLIL* mutation on protein accumulation, we have compared significant spot variations of the proteomes of mature seeds from stop mutants and corresponding wild types. This analysis did not reveal changes in specific spots in mutant seeds. Only few spots showed significant variation, which corresponded to minor modifications in accumulation, insufficient to account for the difference in overall protein content. This experiment highlighted that the increase in storage protein content is probably due to a global increase of spot accumulation in *lil* mutants without affecting a specific spot class. In conclusion, *PsLIL* plays a central role in regulation of seed filling but probably not through direct activation of storage protein gene transcription. This result is consistent with recent studies in *A. thaliana*, which have shown that HAP3 genes





were capable of indirectly activating the expression of seed storage proteins via *ABI3*, *FUS3* and *LEC2* genes (Kagaya *et al.* 2005a; To *et al.* 2006).

#### Role of PsL1L and hypothesis about its potential antagonist role with LEC1

The *LEC1* gene class consists of two sequence groups, the LEC1-type and the L1L-type. Sequence similarity between these groups is restricted to the B domain and not in the A and C domains in *A. thaliana* (Kwong *et al.* 2003) and in *M. truncatula* (unpublished results). Loss of function mutations in either *L1L* or *LEC1* cause defects in embryo development but induce different embryonic phenotypes. *l1l* mutants arrest embryogenesis early in their development, during globular stages (Kwong *et al.* 2003), whereas, *lec1* mutants arrest at a later stage of embryogenesis and exhibit defects in desiccation processes (for review Harada *et al.* 2001). Thus LEC1 and L1L have distinct endogenous functions, a finding which is supported by other observations. First, their expression profiles are different with *LEC1* RNA accumulating earlier in embryogenesis than *L1L* RNA (Kwong *et al.* 2003 and Verdier, unpublished results). Secondly, *L1L* RNA detected in *lec1* mutants is insufficient to prevent defects induced by *lec1* mutation (Kwong *et al.* 2003). Despite these differences, studies have shown that the *L1L* gene could partially rescue an *A. thaliana lec1* phenotype when expressed under the control of DNA sequences flanking *LEC1* (Kwong *et al.* 2003 and Alemanno *et al.* 2008). This observation suggests a critical role of B domain for their function in embryogenesis.

Concerning the process of seed filling, *lec1* mutants have lower seed storage protein content (West *et al.* 1994 and Kwong *et al.* 2003), consistent with a role for *AtLEC1* in seed storage protein regulation. The *AtLEC1* gene controls the expression of seed storage protein genes in a hierarchical manner, which involves *ABI3* and *FUS3* (Kagaya *et al.* 2005a). However in our study, the putative loss of function of *PsL1L* results in seeds with higher seed storage protein content, suggesting a negative regulatory role of this gene on storage protein synthesis. How can one reconcile the finding that *L1L* expressed from *LEC1* gene-flanking sequences is able to complement a *lec1* embryo-defective phenotype with the proposal that L1L and LEC1 may have antagonistic roles during seed filling?

There are several potential explanations, already advanced by Kwong *et al.* (2003), but two of them appear more likely: the sites of gene expression are critical for their functions, or L1L is a regulator of *LEC1* expression. In the first hypothesis, LEC1 and L1L proteins are functionally equivalent during embryogenesis but their role in seed storage synthesis initiation depends on their cellular localisation. In the second hypothesis, L1L would be negative regulator of LEC1 expression. Both hypotheses are consistent with the study of To *et al.* (2006), which established that gene function varies spatially throughout the embryo implying a network of local and redundant gene regulation between the four major regulators of seed maturation (*LEC1*, *LEC2*, *ABI3* and *FUS3*). Additional information is needed to distinguish between these possibilities.

In conclusion, we have identified a HAP3 subunit closely related to L1L in pea. This gene expressed at the transition between embryogenesis and seed filling. Mutant analyses have shown a negative influence of *PsL1L*



on storage protein accumulation in mature seed. Thus, *PsL1L* could represent a negative regulator of seed storage protein synthesis in pea.

## SUPPLEMENTARY MATERIAL

Figure S1: 2-DE analysis of pea *l1l* mutants.

(a) 2-DE gel of pea mature seed with spots up-accumulated in STOP mutants compared to wild-types in red and down-accumulated in green. Spot accession numbers are indicated to refer to Table S2.

(b) Pie chart representing relative abundance with respect to percentage of total volume of up and down-accumulated spots. The number of spots present in each class is indicated in brackets.

Table S1: List of genes profiled in qRT-PCR with annotations, primer sequences and experimental data.

Table S2: List of proteins differentially accumulated between mutants and corresponding wild-types and their relative abundance.

## ACKNOWLEDGMENTS

We are grateful to all members of laboratory INRA UMR-LEG, particularly Grégoire Aubert and Vincent Savoie for their helpful discussions; Michaël Bourgeois for his very valuable advices and his proteomic expertise and Florence Cassecuelle for useful assistance and availability with FPLC analyses.

## REFERENCES

- Alemanno, L., Devic, M., Niemenak, N., Sanier, C., Guillemot, J., Rio, M., Verdeil, J.L., Montoro, P. (2008) Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. *Planta* 227(4):853-66.
- Allen SE (1974) *Chemical analysis of ecological materials*. New York : John Wiley and Sons.
- Baniel, A., Bertrand, D., Lelion, A., Gue' guen, J., 1998. Variability in protein composition of pea seed studied by FPLC and multidimensional analysis. *Crop Science* 38, 1568–1575.
- Bhattacharyya MK, Smith AM, Ellis THN, Hedley C and Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*, 60(1): 115-122.
- Boulter, D. (University of Durham (UK). Department of Botany); Evans, I.A.; Ellis, J.R.; Shirsat, A.; Gatehouse, J.A.; Croy, R.R.D. (1987) Differential gene expression in the development of *Pisum sativum*



[transcriptional control, post-transcriptional control, legumin, vicilin] *Plant Physiology and Biochemistry*. 25(3) p. 283-289

Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 1976 May 7;72:248-54.

Cernac A, Benning C. (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. *Plant J*. 2004 Nov;40(4):575 -85.

Chang S, Pur Year J, Carney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep* 11:113-116

Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savoies V, Aubert G, Brunaud V, de Oliveira Y, Guichard C, Thompson R, Bendahmane A. (2008) UTILdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol*. 2008;9(2):R43.

Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savoies V, Aubert G, Brunaud V, de Oliveira Y, Guichard C, Thompson R, Bendahmane A. (2008). UTILdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol*. 2008;9(2):R43.

Edwards, D., Murray, J.A., Smith, A.G. (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in *Arabidopsis*. *Plant Physiol*. 117(3):1015-22.

Fambrini M, Durante C, Cionini G, Geri C, Giorgetti L, Michelotti V, Salvini M, Pugliesi C. (2006) Characterization of LEAFY COTYLEDON1-LIKE gene in *Helianthus annuus* and its relationship with zygotic and somatic embryogenesis. *Dev Genes Evol*. 216(5):253-64.

Fambrini M, Durante C, Cionini G, Geri C, Giorgetti L, Michelotti V, Salvini M, Pugliesi C. Characterization of LEAFY COTYLEDON1-LIKE gene in *Helianthus annuus* and its relationship with zygotic and somatic embryogenesis. *Dev Genes Evol*. 2006 May;216(5):253-64.

Finkelstein RR, Lynch TJ. (2000) The *Arabidopsis* abscisic acid response gene *ABI5* encodes a basic leucine zipper transcription factor. *Plant Cell*. 2000 Apr;12(4):599-609.

Gallardo K, Firnhaber C, Zuber H, Hélicher D, Belghazi M, Henry C, Küster H, Thompson R. (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol Cell Proteomics*. 2007 Dec;6(12):2165-79.

Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D. (2002). Proteomics of *Arabidopsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol*. 129(2):823-37.



- Garcia D, Fitz Gerald JN, Berger F. (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell*. 2005 Jan;17(1):52-60.
- Gatehouse JA, Evans IM, Bown D, Croy RR, Boulter D. (1982) Control of storage-protein synthesis during seed development in pea (*Pisum sativum* L.) *Biochem J*. 1982 Oct 15; 208(1):119-27.
- Graham PH, Vance CP. (2003) Legumes: importance and constraints to greater use. *Plant Physiol*. 2003 Mar;131(3):872-7.
- Gusmaroli G, Tonelli C, Mantovani R. Regulation of the CCAAT-Binding NF-Y subunits in *Arabidopsis thaliana*. *Gene*. 2001 Feb 21;264(2):173-85.
- Harada JJ (2001) Role of Arabidopsis LEAFY COTYLEDON genes in seed development. *Journal of plant Physiology*, vol. 158, no 4 pp. 405-409
- Hardtke CS, Berleth T. (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryonic axis formation and vascular development. *EMBO J*. 1998 Mar 2;17(5):1405-11.
- Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T. (2005) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol*. 2005 Mar;46(3):399-406.
- Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., Yamamoto, A., Hattori, T. (2005a) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. *Plant Cell Physiol*. 46(2):300-11.
- Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., Hattori, T. (2005b) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol*. 46(3):399-406.
- Kim IS, Sinha S, de Crombrughe B, Maity SN. Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol*. 1996 Aug;16(8):4003-13.
- Künne C, Lange M, Funke T, Miehe H, Thiel T, Grosse I, Scholz U. CR-EST: a resource for crop ESTs. *Nucleic Acids Res*. 2005 Jan 1;33(Database issue):D619-21.
- Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ. (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell*. 2003 Jan;15(1):5-18.

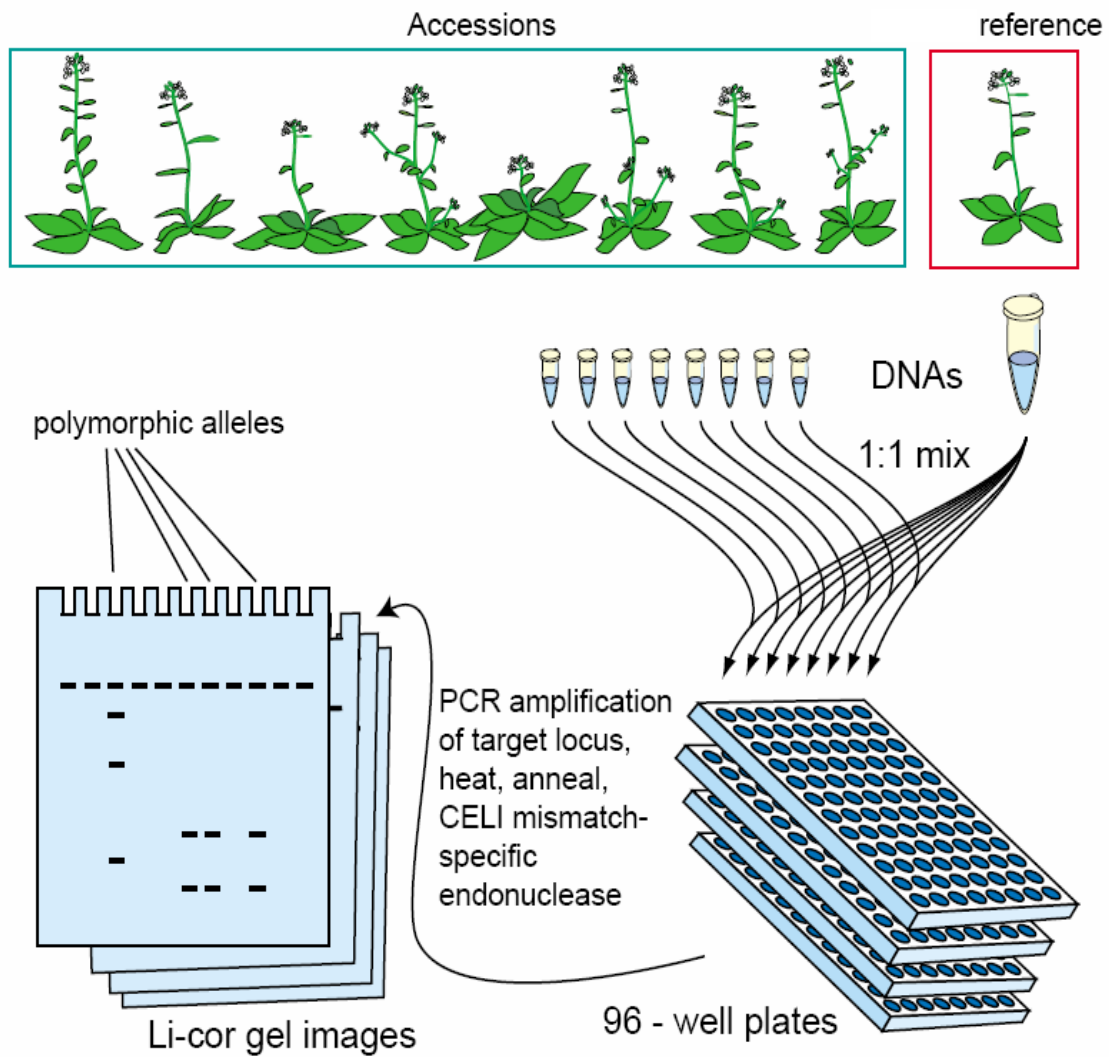




- Ladin, B.F., Tierney, M.L., Meinke, D.W., Hosángadi, P., Veith, M., and Beachy, R.N. (1987) Developmental Regulation of  $\beta$ -Conglycinin in Soybean Axes and Cotyledons. *Plant Physiol.* 84(1): 35–41.
- Lee H, Fischer RL, Goldberg RB, Harada JJ. (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci U S A.* 2003 Feb 18;100(4):2152-6.
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell.* 1998 Jun 26;93(7):1195-205.
- Mathesius, U., Keijzers, G., Natera, S.H., Weinmann, J.J., Djordjevic, M.A., and Rolfe, B.G. (2001) Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics*, 1, 1424-1440
- McCallum CM, Comai L, Greene EA, Henikoff S. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 2000 Jun;123(2):439-42.
- Meinke DW, Franzmann LH, Nickle TC, Yeung EC. (1994) Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell.* 1994 Aug;6(8):1049-1064.
- Meinke, D.W., Chen, J. and Beachy, R.N. (1981) Expression of storage-protein genes during soybean seed development. *Planta* 153(2): 130-139.
- Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003 Jul 1;31(13):3812-4.
- Pang, P.P., Pruitt, R.E. and Meyerowitz, E.M. (1988) Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of *Arabidopsis thaliana*. *Plant Mol. Biol.* 11(6): 805-820.
- Parcy, F., Valon, C., Kohara, A., Miséra, S., Giraudat, J. (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell.* 9(8):1265-77.
- Shujun Chang<sup>1</sup>, Jeff Puryear<sup>1</sup> and John Cairney (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* Volume 11(2): 113-116
- Sinha S, Kim IS, Sohn KY, de Crombrughe B, Maity SN. Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol.* 1996 Jan;16(1):328-37.



- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994 Nov 11;22(22):4673-80.
- To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., Parcy, F. (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* 18(7):1642-51.
- Unger E, Parsons RL, Schmidt RJ, Bowen B, Roth BA. (1993) Dominant Negative Mutants of Opaque2 Suppress Transactivation of a 22-kD Zein Promoter by Opaque2 in Maize Endosperm Cells. *Plant Cell.* 1993 Aug;5(8):831-841.
- Verdier J, Kakar K, Gallardo K, Le Signor C, Aubert G, Schlereth A, Town CD, Udvardi MK, Thompson RD. (2008) Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol Biol.* 2008 Jun 5.
- Walling, L., Drews, G.N. and Goldberg, R.B. (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci U S A.* 83(7): 2123–2127.
- Wang TL, Domoney C, Hedley CL, Casey R, Grusak MA (2003) Can we improve the nutritional quality of legume seeds? *Plant Physiol.* 2003 Mar;131(3):886-91.
- West M, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ. (1994) LEAFY COTYLEDON1 is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in Arabidopsis. *Plant Cell.* 1994 Dec;6(12):1731-1745.
- Yazawa K, Takahata K, Kamada H. (2004) Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol Biochem.* 2004 Mar;42(3):215-23.



**Figure 4.1:** Principe de l'EcoTILLING, schéma tiré de l'article fondateur de cette méthode (Comai et al., 2004)

### 3. ETUDE DU POLYMORPHISME DE SEQUENCE DE *PsL1L* DANS UNE COLLECTION DE 372 ECOTYPES CONTRASTES DE POIS

La capacité d'une espèce à s'adapter à son environnement réside dans sa variabilité génétique. Cette variabilité peut se manifester sous plusieurs formes dont l'une des plus fréquentes est le changement d'un simple nucléotide (SNP, Single Nucléotide Polymorphism). Les SNP participent aux stratégies adaptatives de l'histoire des populations et se sont accumulés durant l'évolution. Par ailleurs, ce polymorphisme naturel du génome a souvent été réduit lors des différentes étapes de domestication des espèces agronomiques. A l'heure actuelle, de nombreuses études utilisent cette diversité car elle représente un atout pour l'élucidation de la fonction des gènes.

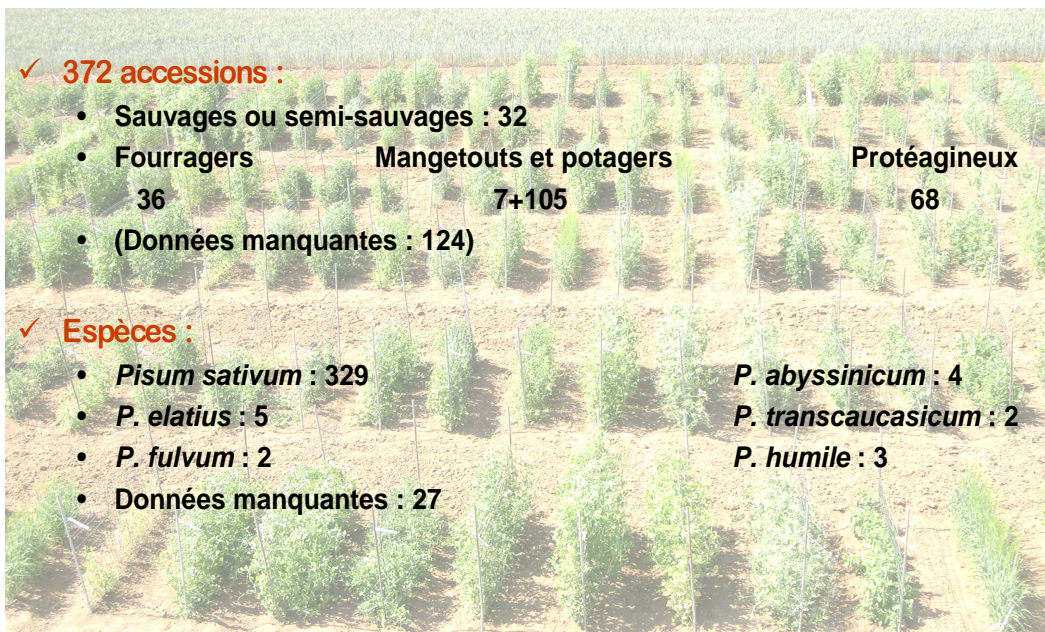
#### **3.1 Choix de la méthode d'EcoTILLING et présentation de la collection CRB de pois**

##### **3.1.1 Méthode d'EcoTILLING (Comai *et al.* 2004)**

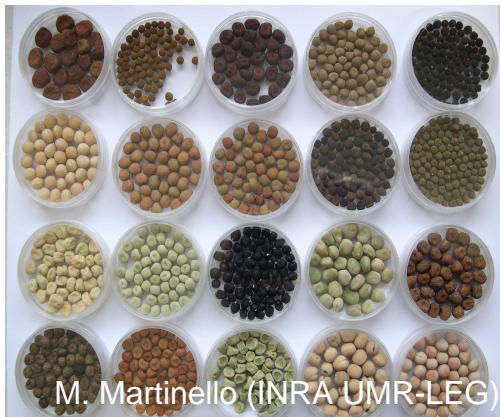
Plusieurs techniques existent pour caractériser la variabilité d'une séquence nucléique mais toutes comportent de nombreux inconvénients : les techniques d'électrophorèse ou de chromatographie de mobilité différentielle tels que le Denaturing Gradient Gel Electrophoresis (DGGE) et le Single Strand Conformation Polymorphism (SSCP) ne permettent pas d'obtenir une position précise du polymorphisme, ni le type de polymorphisme identifié (de Francesco et Perkel, 2001); les techniques récentes de High Resolution Melting (HRM) basées sur les cinétiques de dénaturation des fragments ADN obtenues en PCR quantitative comportent de nombreuses restrictions techniques telles que la taille des fragments à étudier (Gundry *et al.* 2003) ; les techniques basées sur les hybridations de puce à ADN sont coûteuses et peu sensibles avec seulement 50% des SNPs identifiés. Enfin le séquençage haut débit paraît être la solution idéale dans ce genre d'analyse mais est encore très coûteux même si l'on constate une baisse importante du coût de ces analyses depuis quelques années avec l'arrivée de nouvelles technologies (454, Illumina, SOLiD, ...). Le TILLING (Targeted Induced Lesions In Genomes, précédemment décrit chapitre 2 section 4.1.1.1) permet d'identifier le polymorphisme d'une séquence génomique dans un fragment d'intérêt d'environ 1kb (McCallum *et al.* 2000a, b et Colbert *et al.* 2001). Cette technique, peu coûteuse quand elle est couplée à une détection sur gel de séquence, repose sur une amplification PCR du fragment à étudier à l'aide d'amorces marquées puis d'une digestion enzymatique avec une endonucléase qui coupe au niveau des mésappariements, la localisation du mésappariement étant ensuite identifiée sur des gels de séquençage grâce à la taille des produits de digestion.

Cette technique, qui a été développée pour identifier des mutations au sein d'une population de plantes mutantes traitées à l'EMS, a été adaptée pour identifier le polymorphisme naturel de séquence entre une collection de plusieurs écotypes et un écotype de référence, et a été appelée EcoTILLING (voir Figure 4.1, Comai *et al.* 2004).

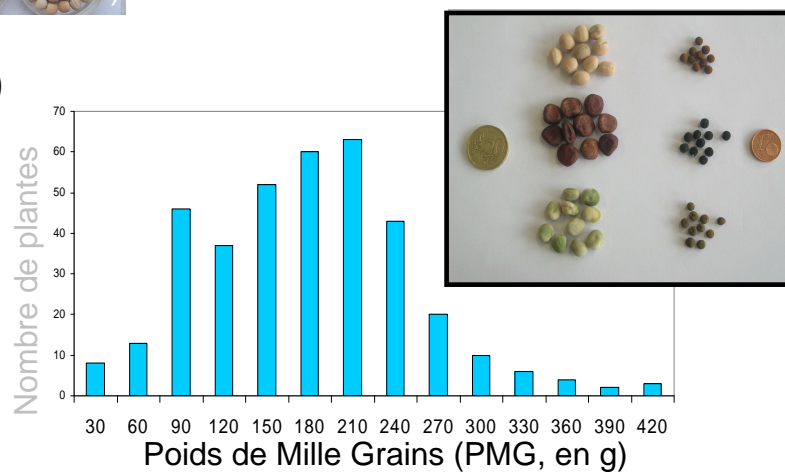
(a)



(b)



(c)



**Figure 4.2 : Présentation de la collection de référence de pois du CRB INRA UMR-LEG**

- (a) Description générale des 372 écotypes
- (b) Diversité phénotypique des graines de la collection de référence
- (c) Représentation de la distribution du caractère « poids de mille grains » dans la collection de référence. Photo représentant la variabilité de la taille d'une graine dans cette collection



L'EcoTILLING permet donc de mettre en évidence du polymorphisme naturel de séquence entre une collection d'écotypes et ce génotype de référence.

### 3.1.2 Collection CRB d'écotypes de pois

Dans cette étude, nous disposons de la collection de 372 écotypes de pois maintenue au sein de l'unité par le CRB INRA UMR-LEG (figure 4.2a, H. Charrel, JB Magnin Robert, J. Burstin et G. Duc, publication en cours). Cette collection de référence a été construite à partir d'environ 1700 accessions de pois répertoriées par le CRB INRA UMR-LEG. Parmi toutes ces accessions, 372 écotypes ont été sélectionnés à partir de données phénotypiques et moléculaires (marqueurs microsatellites) dans le but de constituer une collection de référence la plus contrastée possible. Cette collection regroupe six espèces de pois du monde entier. Elle est classifiée en fonction des types variétaux (fourragers, protéagineux, espèces sauvages, ...) et 43 caractères phénotypiques ont été décrits et suivis pour chacune des 372 accessions. Cette collection présente en particulier une grande variabilité pour de nombreux caractères associés à la graine, comme par exemple la pigmentation (figure 4.2b) ou le poids de mille grains (PMG) (valeurs allant de 30 à 420 g en fonction des écotypes, figure 4.2c).

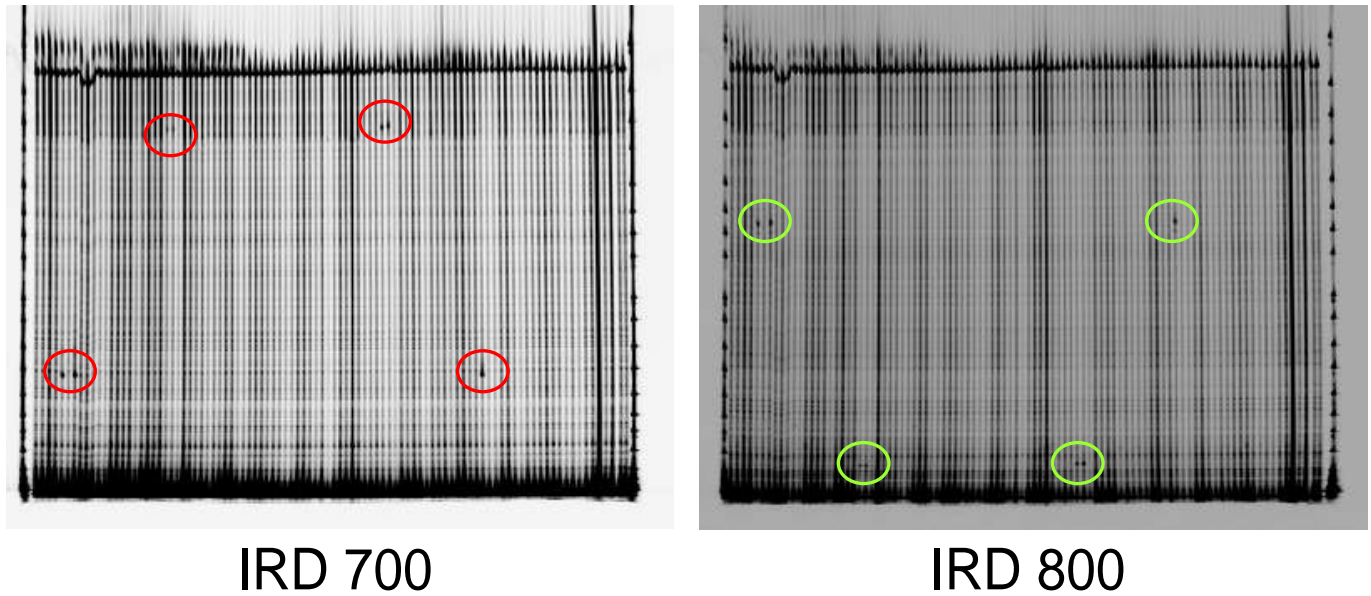
Parmi les 372 accessions, nous avons choisi d'utiliser l'écotype Caméor comme écotype de référence pour cette analyse. Cet écotype correspond à un ancien cultivar de pois de conserve, d'aspect feuillu à entre-nœuds courts et à teneur élevée en protéines (approximativement 24% de la graine mature). Il est couramment utilisé comme référence au laboratoire et par conséquent de nombreuses données génomiques ont été accumulées sur ce génotype. C'est également le génotype de référence utilisé pour la banque BAC (INRA URGV) et pour la population de mutants EMS/TILLING (Collaboration INRA URGV et UMR-LEG). Il a également été utilisé comme parent dans de nombreux croisements ayant généré des populations de cartographie génétique.

### 3.2 Analyse du polymorphisme de séquence de *PsL1L*

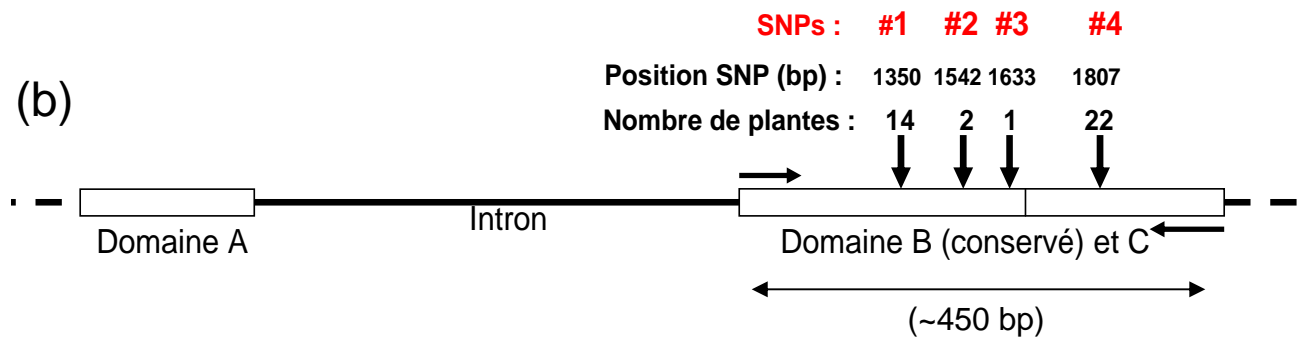
Sur cette collection de référence, la détection du polymorphisme naturel de séquence a été réalisée par EcoTILLING sur un fragment du gène *PsL1L* d'environ 450 pb. Ce fragment d'étude correspond au deuxième exon du gène contenant les domaines B et C (voir figure 4.3) : le domaine B très conservé parmi les gènes HAP3 et le domaine C spécifique du gène *PsL1L*. L'amplification PCR de ce fragment a été réalisée sur l'ADN génomique des 372 écotypes de pois. Ce produit d'amplification a été mélangé avec celui correspondant à ce même fragment amplifié sur le cultivar de référence Caméor. Quatre sites de polymorphisme naturel de séquence (SNP) ont ainsi pu être identifiés dans cet exon (Figure 4.2 et Tableau 4.1). Trois sites polymorphes ont été identifiés dans le domaine B (SNP #1, #2 et #3 respectivement en position 1350, 1542 et 1633 pb par rapport à la séquence génomique) alors que seulement un site polymorphe a pu être identifié dans le domaine C (SNP #4 en position 1807 pb). Ces sites sont polymorphes chez certains écotypes et présentent donc un allèle différent de



(a)



(b)



**Figure 4.3 : Détection du polymorphisme de séquence du gène *PsLIL***

- (a) Exemples de gels de séquence LI-COR obtenus aux deux longueurs d'onde d'observation 700 et 800 nm. Chaque ligne de migration représente le mélange entre le fragment PCR amplifié d'un écotyle de la collection et celui de l'écotype de référence, Caméor. Les produits clivés par l'endonucléase sont entourés sur les gels et les tailles de ces produits correspondent à la position du polymorphisme. Pour que la détection du polymorphisme soit considérée comme fiable, deux bandes complémentaires doivent être identifiées à chacune des deux longueurs d'onde utilisée.
- (b) Représentation schématique de la position des SNP identifiés sur le gène *PsLIL*. La position des exons, introns, domaines et de la région étudiée est indiquée. La position exacte et le nombre d'écotype portant les différents SNP sont également indiqués.

celui de Caméor. L'allèle de "Caméor" aux différents SNP est largement majoritaire. Ainsi la fréquence d'apparition des allèles "minoritaires" est faible : nous avons identifié 14, 2, 1 et 22 écotypes présentant respectivement une variation allélique au niveau des SNP #1, #2, #3 et #4 (figure 4.3, Tableau 4.1).

Ce premier résultat montre que l'haplotype de type "Caméor" est majoritaire et que donc les domaines B et C sont peu polymorphes au sein de cette collection. Cette conservation du gène suggère que ces séquences ont été soumises à sélection. Sur la base de ces résultats, on peut émettre l'hypothèse que ces domaines jouent donc un rôle prépondérant dans la fonction du gène. Ce résultat avait déjà été démontré en ce qui concerne le domaine B qui contient deux régions essentielles à la fonction du gène (les régions de fixation à l'ADN et de fixation des sous-unités) (Lee *et al.* 2003). Cependant, cette constatation est également valable pour le domaine C. Malgré le fait que ce domaine soit peu conservé entre espèces, il apparaît être soumis à une forte pression de sélection chez le pois.

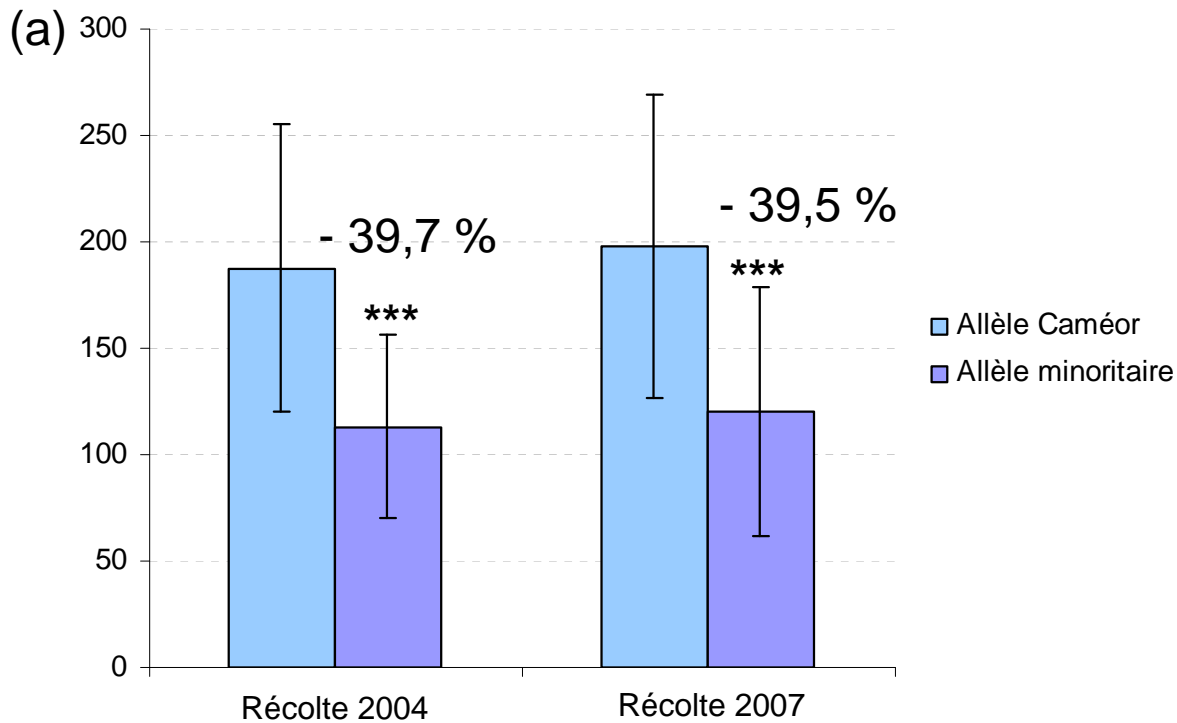
La validité, la position précise et la nature des différents SNP identifiés ont été déterminés par séquençage de certains écotypes portant les différents allèles "minoritaires". Les SNP #2, #3 et #4 ont ainsi été validés alors que le SNP #1 situé dans les 50 premiers nucléotides du fragment séquencé n'a pas pu être validé en raison de la difficulté d'analyse des chromatogrammes de séquence aux extrémités. Ce résultat confirme le fait que l'EcoTILLING est une méthode robuste d'identification des SNP car en accord avec des données de séquençage. Une information supplémentaire est cependant apparue dans les données de séquence et n'a pas pu être mise en évidence dans l'analyse EcoTILLING : le SNP #4 semble être associé à un autre SNP distant de huit nucléotides sur l'échantillonnage des quelques écotypes séquencés (Annexe 3a).

Les séquences nucléotidiques obtenues ont été traduites *in silico* pour vérifier si les séquences protéiques des allèles "minoritaires" étaient modifiées au niveau des SNP identifiés. Sur les trois SNP séquencés, deux (SNP #3 et #4) ont été associés à des variations silencieuses et n'entraînent donc aucune modification dans la séquence protéique par rapport à l'allèle "Caméor". Seul l'allèle "minoritaire" du SNP 2 entraîne une modification de la séquence protéique avec la substitution d'une thréonine (T) par une isoleucine (I) (Annexe 3b).

### **3.3 Génétique d'association entre le SNP #4 et des caractères de qualité de la graine**

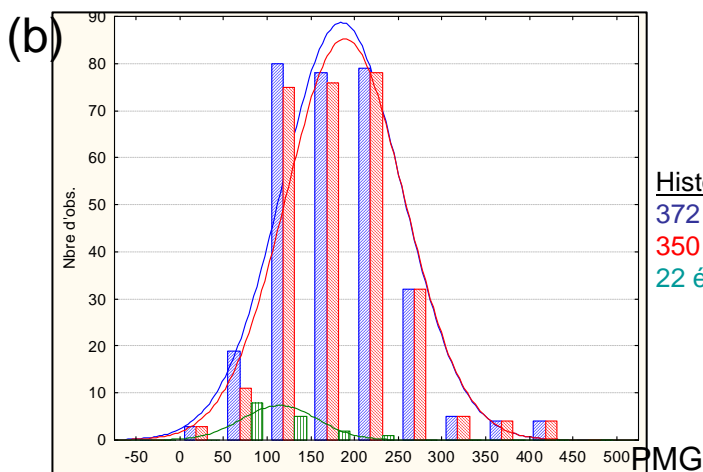
#### **3.3.1 Génétique d'association**

Comme décrit précédemment, les écotypes appartenant à la collection de référence du CRB ont été phénotypés au champ, sur deux années différentes, pour 43 caractères phénotypiques allant de l'architecture de la plante, à des caractères associés aux fleurs, aux graines mais également à la productivité et au rendement. Des tests statistiques ont été réalisés (ANOVA à un facteur) afin de mettre en évidence d'éventuelles corrélations entre le polymorphisme naturel de séquence précédemment identifié chez les plantes portant les SNP #1 et #4 et des caractères liés à la qualité du grain. Les SNP #2 et #3 n'ont pas pu être analysés car trop peu représentés dans la collection (respectivement deux et un individu présentant ces SNP). Pour le SNP #1, aucune corrélation n'a pu



Récolte 2004	PMG (mg)	N	Ecart-type
Allele « Caméor »	187,68	288	67,38
Allèle « minoritaire »	113,19	16	43,36

Récolte 2007	PMG (mg)	N	Ecart-type
Allele « Caméor »	198,18	343	71,25
Allèle « minoritaire »	119,87	22	58,47



Histogramme catégorisé du PMG de :  
 372 écotype de la collection  
 350 écotype « allèle Caméor »  
 22 écotypes « allèle minoritaire »

**Figure 4.4 : Corrélation entre le PMG et les différents allèle du SNP #4**

- (a) Graphique représentant la valeur du PMG en fonction de l'allèle du SNP #4 et de l'année de récolte.
- (b) Histogramme catégorisé du PMG en fonction du nombre d'écotype. En bleu, représentativité de la collection complète de 372 écotypes, en rouge les écotypes portant l'allèle Caméor, en vert les écotypes portant l'allèle minoritaire

être mise en évidence entre ce polymorphisme et les caractères décrits de la collection. En revanche, une très forte corrélation statistique a été mise en évidence entre le SNP #4 et le poids de mille grains (PMG) des plantes. En effet, on constate que les plantes identifiées avec l'allèle "minoritaire" du SNP #4 ont des graines dont le poids de graine mature est 40% plus faible (p-value  $\ll 0.001$ ). Cet effet est d'autant plus intéressant qu'il est très reproductible sur les deux années étudiées : baisse significative de 39,7% (p-value  $< 2.10^{-5}$ ) et 39,5% (p-value  $< 10^{-6}$ ) du PMG respectivement sur les récoltes 2004 et 2007 (figure 4.4). Cet allèle étant associé à un changement silencieux d'acide aminé, nous pouvons donc aboutir à deux hypothèses :

- soit le SNP #4 est directement corrélé à ce caractère agronomique, et est responsable d'une modification de la stabilité de l'ARNm du gène *PsLIL* (Potokina *et al.* 2006),
- soit le SNP #4 est indirectement corrélé avec ce caractère, car associé avec un autre SNP non identifié par déséquilibre de liaison.

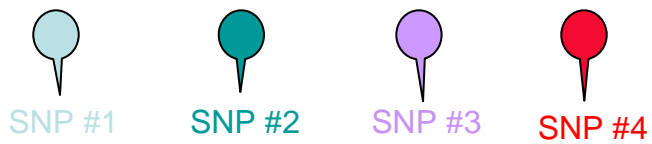
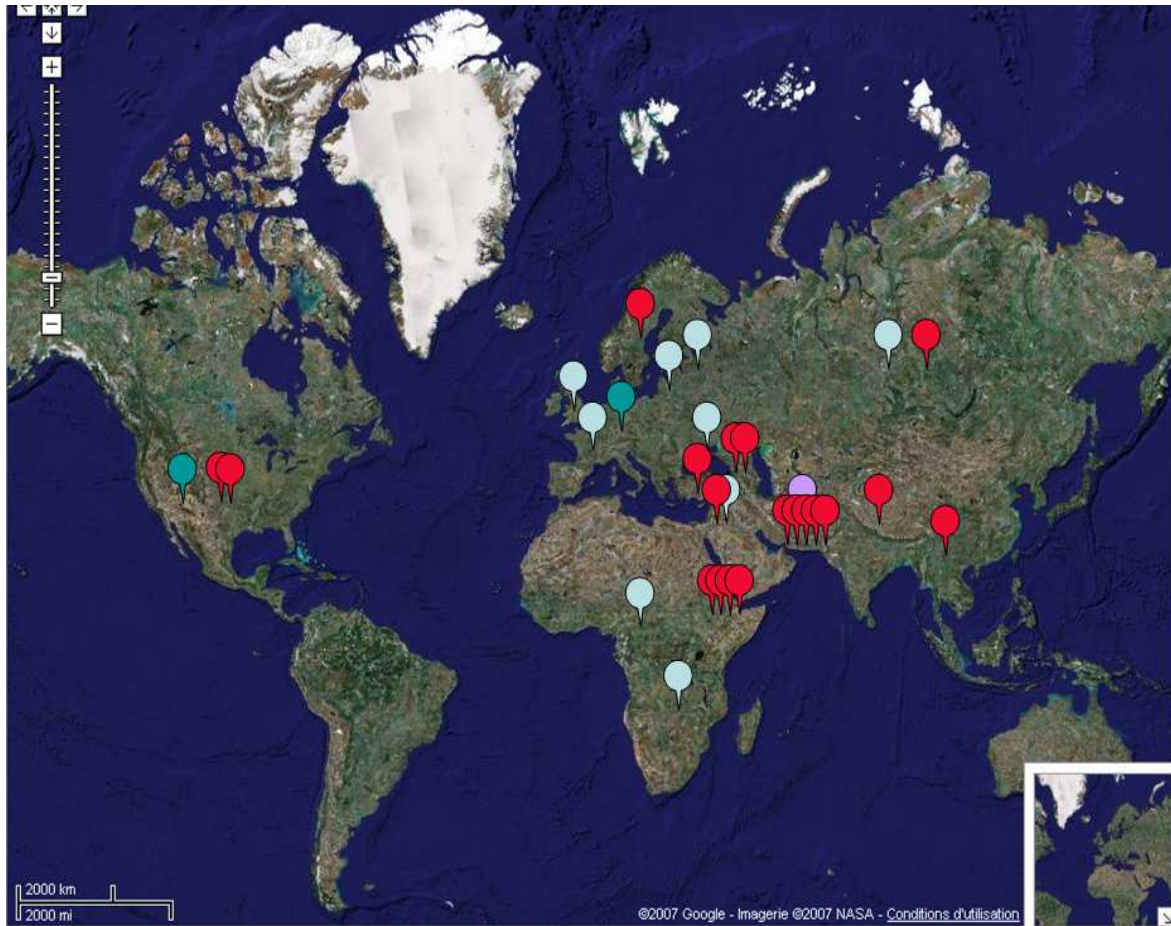
### 3.3.2 Localisation géographique des différents allèles du SNP #4

Grâce aux données passeports des différents écotypes de pois présents dans la collection, nous avons voulu analyser les éventuelles corrélations entre cette variabilité naturelle et la localisation géographique des écotypes portant les différents allèles (Figure 4.5), en particulier ceux portant les allèles "minoritaires" du SNP #4.

L'allèle "minoritaire" du SNP #4 a été retrouvé dans 22 écotypes différents, principalement des variétés de *P. sativum* de type feuillu annotées comme variétés sauvages. En effet, l'ensemble de la collection des 372 écotypes ne comprend qu'environ 8,6% (32/372) de variétés sauvages. A l'inverse, dans notre analyse, environ 45% (10/22) des écotypes portant cet allèle se sont révélés être des variétés sauvages. Cette observation est renforcée par la localisation géographique des écotypes. La collection est composée d'environ 3,5% (13/372) d'écotypes provenant d'Ethiopie et 1,6% (6/372) provenant d'Afghanistan. Lorsque l'on analyse le lien entre la présence de cet allèle "minoritaire" du SNP #4 et la provenance géographique des écotypes, on constate que plus de 66% (4/6) des écotypes d'Afghanistan et que 38% des écotypes d'Ethiopie représentés ont été identifiés dans notre analyse avec cet allèle "minoritaire". Si l'on ajoute le fait que la région d'origine du pois se situe vraisemblablement en Abyssinie (pointe nord de l'Ethiopie) et en Afghanistan, on peut donc conclure que cet allèle "minoritaire" est préférentiellement associé à des variétés sauvages de pois. Nous pouvons donc émettre l'hypothèse selon laquelle cet allèle "minoritaire" serait une forme allélique ancestrale du gène *PsLIL*.

### 3.3.3 Phénotypage et structuration des plantes portant l'allèle "minoritaire" du SNP #4

Suite à la mise en évidence de cette corrélation entre le SNP #4, un caractère agronomique et l'origine sauvage de l'allèle "minoritaire", nous avons analysé les phénotypes des plantes portant l'allèle lors d'une expérimentation au champ pour valider notre hypothèse d'allèle ancestral (Tableau 4.2). Nous avons constaté que



**Figure 4.5 :** Localisation géographique d'origine des écotypes portant les allèles « minoritaires » pour chacun des SNP identifiés

les plantes portant l'allèle "minoritaire" présentaient des phénotypes assez variés : des variations de tailles importantes (de 40 cm à plus d'1,5 m), des densités de feuillage ainsi que des tailles de gousses variables. Le seul caractère phénotypique commun entre ces écotypes était la présence de graines de petites tailles. Cette variabilité des phénotypes associés à cet allèle pourrait suggérer une origine lointaine de cet allèle "minoritaire".

Pour conforter cette analyse phénotypique, nous avons voulu regarder la structuration génétique des écotypes portant cet allèle "minoritaire". La structuration de la collection de référence comportant les 372 écotypes a été analysée en utilisant 29 marqueurs microsatellites et trois marqueurs morphologiques grâce à plusieurs méthodes (algorithme STRUCTURE et clustering WARD ou UPGMA) (J. Burstin, publication en cours). On constate que cette analyse conforte les observations faites au champ, à savoir que les plantes portant l'allèle "minoritaire" ne sont pas regroupées dans un seul et même groupe mais qu'elles apparaissent au contraire dans différents groupes plus ou moins éloignés (figure 4.6). Par exemple, si l'on utilise la classification en 12 clusters de ces écotypes (critère d'agglomération de Ward), on constate que les 22 écotypes portant l'allèle "minoritaire" du SNP #4 sont répartis dans 6 groupes différents.

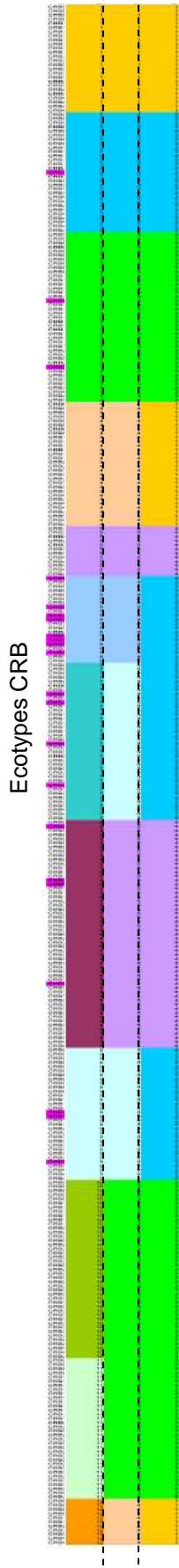
En conclusion, ces observations phénotypiques et moléculaires pourraient suggérer que les plantes portant l'allèle "Caméor" ont été sélectionnées au cours de la domestication de manière convergente, au sein de plusieurs groupes distincts révélés par le cluster. La présence de plante portant l'allèle "minoritaire" dans plusieurs groupes de structuration de la collection, pourrait correspondre au fait que certains écotypes n'ont probablement pas été soumis à sélection pour le caractère "Poids de mille graines".

### **3.4 Implication de l'allèle lié au SNP #4 dans l'amélioration d'un caractère agronomique du pois ?**

Cette étude a permis de mettre en évidence un allèle "minoritaire" pour un SNP présent dans le gène *PsL1L*, allèle qui a été identifié préférentiellement dans des variétés sauvages de pois. Cette dernière observation semble cohérente avec la localisation géographique des plantes portant l'allèle "minoritaire" qui sont pour plus de la moitié originaires d'Afghanistan et d'Ethiopie, régions d'origine probable du pois. Cet allèle est fortement corrélé à un caractère agronomique : les plantes portant l'allèle "Caméor" présentent une augmentation du poids de la graine de 40% par rapport à celles portant l'allèle "minoritaire". Cette corrélation négative entre la présence de l'allèle "minoritaire" et un caractère agronomique de la graine explique que cet allèle soit peu représenté dans la collection de référence CRB qui est majoritairement composée de cultivars. En conclusion, nous pouvons émettre l'hypothèse que cet allèle pourrait être un allèle ancestral, présent essentiellement dans des variétés de pois sauvage à faible poids de grains, qui aurait été contre-sélectionné durant la domestication du pois, domestication dont un des objectifs a été naturellement d'augmenter la taille et donc le poids de la graine. Ainsi, l'allèle "Caméor" aurait été soumis à une sélection convergente dans de nombreuses populations cultivées pour son effet positif sur le PMG, permettant une augmentation de 40% du poids du grain chez le pois.



(a)  
 Cluster WARD  
 Nombre de clusters : 12 7 4



Structuration (STRUCTURE)  
 Classification WARD  
 Classification UPGMA

(b)

Ecotypes portant l'allèle "minoritaire"	_NAME_		Structuration (STRUCTURE)					Classification WARD					Classification UPGMA			
	group4_32	group4_29	CLUSTER_ward_12	CLUSTER_ward_10	CLUSTER_ward_7	CLUSTER_ward_6	CLUSTER_ward_5	CLUSTER_ward_4	CLUSTER_upgma_12	CLUSTER_upgma_10	CLUSTER_upgma_7	CLUSTER_upgma_6	CLUSTER_upgma_5	CLUSTER_upgma_4		
CRB005	3	3	8	8	5	5	4	4	7	7	6	6	5	4		
CRB015	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB081	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB123	3	3	7	7	7	6	5	2	7	7	6	6	5	4		
CRB128	3	3	7	7	7	6	5	2	8	6	5	5	4	2		
CRB135	2	2	3	3	3	3	3	3	3	3	3	3	3	3		
CRB137	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB149	3	2	8	8	5	5	4	4	7	7	6	6	5	4		
CRB150	2	2	8	8	5	5	4	4	7	7	6	6	5	4		
CRB151	3	3	9	9	6	6	5	2	6	6	5	5	4	2		
CRB152	3	3	9	9	6	6	5	2	6	6	5	5	4	2		
CRB194	3	3	2	2	2	2	2	2	2	2	2	2	2	2		
CRB201	3	3	7	7	7	6	5	2	8	6	5	5	4	2		
CRB202	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB294	2	2	3	3	3	3	3	3	3	3	3	3	3	3		
CRB345	3	2	8	8	5	5	4	4	7	7	6	6	5	4		
CRB350	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB351	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB361	3	3	7	7	7	6	5	2	8	6	5	5	4	2		
CRB381	3	3	6	6	6	6	5	2	6	6	5	5	4	2		
CRB394	3	3	9	9	6	6	5	2	6	6	5	5	4	2		
CRB399	3	3	6	6	6	6	5	2	6	6	5	5	4	2		

**Figure 4.6 : Structuration des écotypes portant l'allèle minoritaire du SNP #4**

- (a) Représentation de la structuration de la population de référence CRB (critère d'agglomération Ward) en 12, 7 et 4 clusters différents. Les individus portant l'allèle minoritaire du SNP #4 sont indiqués en violet à gauche
- (b) Représentation détaillée de l'appartenance des 22 écotypes portant l'allèle "minoritaire" en utilisant trois méthodes de structuration : avec le logiciel Structure en utilisant 29 ou 32 marqueurs, avec le critère d'agglomération Ward (en fixant le nombre de clusters de 4 à 12) et enfin avec le critère d'agglomération UPGMA (de 4 à 12 clusters).

#### 4. CONCLUSION DU CHAPITRE

De nombreuses données de génomique, transcriptomique, protéomique et métabolomique sont actuellement générées sur les quelques plantes choisies comme plantes modèles. Cependant, peu de ces données sont transférées et validées sur les espèces cibles. Dans les chapitres précédents, nous avons identifié des facteurs de transcription potentiellement importants pour le développement de la graine chez *M. truncatula*. Une analyse fonctionnelle a été entreprise sur l'un d'entre eux, le gène *MtLIL*, et a révélé qu'il pourrait intervenir dans des aspects de développement de la graine tels que la synthèse des protéines de réserve. Nous avons donc voulu valider le rôle du facteur de transcription orthologue au gène *MtLIL*, dans l'élaboration de la qualité de la graine chez une légumineuse d'intérêt agronomique, le pois.

##### Synthèse des résultats

Cette étude a permis dans un premier temps d'identifier le facteur de transcription *LEAFY COTYLEDON1-LIKE* chez le pois. Ce gène est exprimé au début du remplissage de la graine et pourrait être indirectement impliqué dans la synthèse des protéines de réserve des graines. Deux mutants III identifiés par TILLING chez le pois montrent que les graines des plantes portant la mutation présentent une augmentation de leur contenu protéique à maturité. Des analyses biochimiques fines ont révélé que cette augmentation n'était pas due à une accumulation de protéines spécifiques mais plutôt à une augmentation globale des protéines de réserve et notamment des légumine. Cette observation est cohérente avec l'hypothèse d'un rôle de *PsLIL* de régulation indirecte de la synthèse des protéines de réserve par effet antagoniste avec *PsLECI*. De plus, nous avons constaté que pour une des mutations, celle entraînant l'apparition d'un codon stop, les graines matures présentaient un poids de graine inférieur à celui des graines sauvages. Ce résultat nous montre un possible rôle de *LIL* dans le déterminisme du poids et donc de la taille des graines.

Dans un deuxième temps, nous avons identifié au sein d'une collection de référence d'écotypes de pois, le polymorphisme de séquence associé au gène *PsLIL*. Nous avons ainsi mis en évidence la présence de quatre sites polymorphes dans les domaines B et C. L'un des SNP, le SNP #4, s'est révélé fortement corrélé avec un caractère d'intérêt agronomique, le poids de mille grains. En effet, l'allèle "Caméor" de ce SNP est corrélé avec une augmentation du poids du grain d'environ 40% par rapport à l'allèle "minoritaire", principalement identifié chez des espèces sauvages, qui pourrait être un allèle ancestral de pois.

Ces deux résultats appuient l'hypothèse du rôle de *PsLIL* dans l'élaboration de la qualité de la graine. En effet ce gène est impliqué dans la régulation de l'accumulation des réserves protéiques, mais également dans le déterminisme du poids de la graine mature.





### **Perspectives de ce travail**

Ce travail a permis de valider l'implication du gène *PsL1L* dans des caractères de qualité de la graine chez le pois : l'analyse EcoTILLING de la collection de référence a permis de mettre en évidence un site polymorphe silencieux dans le domaine C du gène *PsL1L* mais qui s'est révélé corrélér à un caractère agronomique intéressant. La question est maintenant de déterminer si cette variabilité du caractère "PMG" est liée directement au gène *PsL1L*.

En effet, plusieurs cas sont possibles :

La première hypothèse repose sur le fait que le gène *PsL1L* est directement responsable de la variabilité du caractère "PMG". Cette corrélation serait donc soit directement liée au SNP #4 présent dans le domaine C ou soit à un autre site de polymorphisme associé au SNP précédent par déséquilibre de liaison. Cet autre site pourrait être situé dans les régions 5' ou 3' non traduites, dans la région intronique ou dans la région promotrice du gène *PsL1L*.

Pour valider cette hypothèse, une analyse de suivi du niveau d'expression du gène doit être réalisée dans les écotypes ayant l'allèle "Caméor" et ceux ayant l'allèle "minoritaire". Si l'on constate une perte de stabilité des ARNm ou une différence du niveau d'expression de ce gène entre les écotypes portant les deux allèles alors il est probable que le gène *PsL1L* soit directement impliqué dans le caractère "PMG". Il resterait alors à valider tout d'abord la nature du polymorphisme responsable du caractère en séquençant les autres régions citées précédemment. Enfin, une validation de l'effet du polymorphisme devra être réalisée par transformation.

En revanche, si aucune différence d'expression du gène *PsL1L* n'était mise en évidence entre les écotypes portant les deux allèles alors la deuxième hypothèse est que ce gène n'est pas le gène responsable de cette variabilité du PMG. Dans ce cas le polymorphisme associé à notre SNP peut se situer dans une zone plus ou moins éloignée du gène *PsL1L*, potentiellement identifiable par clonage positionnel.

En parallèle à ce travail, il serait également intéressant de comparer le contenu en carbone et azote des écotypes portant les deux allèles, afin de déterminer si la baisse du poids de la graine affecte plus le contenu en azote ou en carbone, ou les deux.



# **CONCLUSION GÉNÉRALE**



# Chapitre 5 :

## Conclusion générale

Les résultats acquis au cours de cette étude s'inscrivent dans un cadre général de compréhension des mécanismes de régulation transcriptionnelle, mécanismes pouvant affecter la teneur et la composition en protéines des graines de légumineuses.

Ce travail de thèse se décompose en deux grandes parties : une première sur le contrôle transcriptionnel du développement de la graine des légumineuses. Cette partie repose sur des approches de transcriptomique et a permis de définir les processus moléculaires, ainsi que les stades clefs du développement de la graine chez *M. truncatula*, pour ensuite identifier des facteurs de transcription associés à ces processus. Une seconde partie est consacrée à la caractérisation fonctionnelle d'un facteur de transcription, homologue au gène d'*A. thaliana* le gène *LECI-LIKE (LIL)*, dont le rôle dans le développement de la graine et l'impact sur la qualité de la graine ont été analysés chez *M. truncatula* et *P. sativum*.

### Contrôle transcriptionnel du développement de la graine des légumineuses

Le premier objectif de ce travail de thèse a été de mieux comprendre les mécanismes de développement de la graine en explorant son transcriptome. Le récent développement de la puce Affymetrix de *M. truncatula* nous a permis de suivre l'expression de 51k gènes durant le développement de la graine. Grâce à cette analyse, plus de 15000 gènes différentiellement exprimés au cours de ce processus développemental complexe ont pu être identifiés. Ces gènes ont été répartis dans quatre groupes en fonction de leur profil d'expression : ceux associés à l'embryogénèse tardive (10 JAP), ceux associés au début du remplissage de la graine (16 JAP), ceux associés à la fin du remplissage (20-24 JAP) et enfin ceux associés à la dessiccation (36 JAP). La répartition fonctionnelle des gènes présents dans ces quatre groupes s'est révélée très contrastée permettant de distinguer des processus plus spécifiquement associés à certaines phases du développement de la graine. Les principaux processus identifiés sont une forte activité de division cellulaire à 10 JAP, une importante régulation du métabolisme des carbohydrates et de celui des acides aminés à 16 JAP, une activité de signalisation cellulaire à 20-24 JAP et enfin une très forte activité transcriptionnelle à 36 JAP.

Cette analyse a également permis d'identifier les stades clefs de transition à étudier pour une analyse plus approfondie du contrôle transcriptionnel de la graine. Tout d'abord, durant la période marquée par une intense activité de division cellulaire identifiée grâce à cette analyse et allant de 12 à 16 JAP. En effet, durant cette période la graine va basculer d'un processus de morphogénèse marqué par une intense division cellulaire vers un processus de remplissage marqué par une expansion cellulaire importante. Ce basculement dans la phase de



remplissage est marqué par l'arrêt des divisions cellulaires. Le nombre de divisions cellulaires pendant l'embryogenèse définit le nombre de cellules de l'embryon qui sera un caractère déterminant pour la taille et le poids de la graine mature (Munier-Jolain *et al.* 1998). Les facteurs de transcription exprimés pendant cette période pourraient potentiellement influencer le déclenchement du remplissage de manière plus ou moins précoce.

Ensuite, la période de remplissage (16-24 JAP) est également importante pour comprendre les régulations directes d'activation des protéines de réserve qui s'accumulent principalement sous contrôle transcriptionnel. Les facteurs de transcription associés à la synthèse des différentes familles de protéines de réserve pourraient représenter de potentiels régulateurs directs de ces familles protéiques.

La suite de ce travail a été spécifiquement orientée vers la caractérisation de la régulation transcriptionnelle de type gène-spécifique avec l'identification de facteurs de transcription intervenant dans les phases clefs. Le suivi d'expression de 712 facteurs de transcription a donc été réalisé au cours du développement de la graine en utilisant l'outil de PCR quantitative pour assurer une grande sensibilité de détection. La grande majorité des gènes (512/712) sont exprimés durant le développement de la graine, en revanche seulement 169 présentent un profil d'expression significativement contrasté. En fonction de leur profil d'expression nous avons réparti ces 169 gènes dans les différents stades biologiques de la graine : l'embryogenèse tardive (41 TFs), le remplissage de la graine composé de quatre sous-groupes [transition embryogénèse-remplissage (28 TFs), synthèses des vicilines (15 TFs), des légumine K (15 TFs), des légumine A (22 TFs)] et enfin la dessiccation (48 TFs). L'expression dans les trois tissus de la graine, réalisée pour les facteurs de transcription fortement exprimés pendant la phase de remplissage, a permis de sélectionner des gènes potentiellement associés à des fonctions spécifiques de ces tissus. Ainsi, neuf facteurs de transcription candidats ont été sélectionnés pour leur probable implication dans les mécanismes de remplissage de la graine des légumineuses.

L'autre objectif de ce travail était de présenter un modèle de régulation transcriptionnelle chez les légumineuses. En effet de nombreuses études réalisées sur *A. thaliana* ont permis d'identifier des facteurs de transcription impliqués dans la synthèse des protéines de réserve. Cependant, des différences sont constatées entre le remplissage de la graine d'*A. thaliana* et des légumineuses. Tout d'abord, au niveau composition en globulines, *A. thaliana* ne présente pas, ou très peu de globulines 7S (vicilines/convicilines) alors que chez les légumineuses, ces protéines représentent une part importante de la teneur totale en protéines. Par ailleurs, les cinétiques d'accumulation des protéines de réserve sont différentes avec une synthèse synchrone de toutes les protéines de réserve chez *A. thaliana* et asynchrone chez *M. truncatula* et le pois. Malgré ces différences, on constate que chez *M. truncatula*, plusieurs facteurs de transcription identifiés chez *A. thaliana* comme régulateur du développement de la graine sont également identifiés par notre approche, comme par exemple des homologues de *LEC1*, *FUS3*, *ABI4*, *ABI5*. Ce résultat nous indique que malgré des différences de régulation, le développement de la graine des légumineuses partage probablement un cadre commun de régulation transcriptionnel avec celui d'*A. thaliana*. En revanche, d'autres gènes identifiés dans notre approche ne présentent pas d'homologues chez *A. thaliana* et constituent donc de bons candidats pour l'étude de la synthèse





ordonnée des protéines de réserve chez les légumineuses. Parmi les neuf facteurs de transcription candidats potentiellement impliqués dans le développement de la graine, quatre sont des homologues de gènes connus chez *A. thaliana* comme jouant un rôle dans la régulation des mécanismes de la graine. L'analyse fonctionnelle de ces gènes permettra d'identifier les mécanismes communs entre *A. thaliana* et *M. truncatula*. En revanche, cinq autres candidats n'ont soit pas d'homologues chez *A. thaliana* ou soit pas été caractérisés fonctionnellement chez cette espèce. L'analyse fonctionnelle de ces gènes pourrait permettre d'élucider de nouveaux et/ou spécifiques mécanismes de régulation du développement de la graine des légumineuses.

### Rôle du gène LEC1-LIKE chez les légumineuses

Dans un deuxième temps, ce travail de thèse a été focalisé sur le rôle d'un des facteurs de transcription identifiés par transcriptomique. Le choix s'est porté sur les gènes de type LEC1 et en particulier sur le gène *LEC1-LIKE* (*LIL*) pour plusieurs raisons. Tout d'abord, ce gène présentait un profil d'expression associé à la phase de transition entre embryogenèse et remplissage. Ensuite, il avait déjà été identifié chez *A. thaliana* comme impliqué dans le développement de la graine mais son implication dans la régulation de la phase de remplissage restait à explorer. Notre étude a consisté à analyser le rôle de ce gène chez *M. truncatula*, et à valider son implication putative dans l'élaboration de la qualité de la graine de pois (*P. sativum*). Chez ces deux espèces, ce gène a donc été identifié comme exprimé à la transition entre embryogenèse et remplissage avec un maximum d'expression lors du début de la synthèse des protéines de réserve (globulines 7S). Deux phénotypes majeurs sont observables chez les mutants/transformants pour le gène *LIL*, aussi bien chez *M. truncatula* que chez *P. sativum* : une teneur en protéines des graines plus forte et un poids de graine plus faible. Ces deux phénotypes ont rarement été observés sur les mêmes plantes, excepté chez les mutants EMS/TILLING de pois présentant une mutation entraînant l'apparition prématurée d'un codon STOP en début de séquence protéique. Cette constatation suggère que les différentes mutations affectent le développement des graines à différent degrés. Le fait que le poids et la teneur en protéines des graines soient affectés nous renseigne sur le rôle putatif de ce gène. Dans l'hypothèse la plus simple, ces deux phénotypes peuvent en fait être liés à la dérégulation d'un seul mécanisme qui est la transition entre l'embryogénèse et le remplissage. En effet, la taille de la graine, et donc son poids sont déterminés en partie lors de l'embryogenèse par le nombre de cellules de l'embryon qui correspond au potentiel de croissance de la graine. D'autre part, la synthèse des protéines de réserve est soumise à de nombreuses régulations transcriptionnelles avec, en particulier, le gène *LEC1*, très proche de *LIL*, qui est impliqué dans l'activation des protéines de réserve par l'intermédiaire des gènes *FUS3* et *ABI3* (Kagaya *et al.* 2005a). Le gène *MtLIL* est exprimé durant la phase de transition entre embryogénèse et remplissage, il pourrait donc être impliqué dans la régulation de ce mécanisme. De plus, nous montrons que le gène *MtLIL* a un effet négatif sur la synthèse des protéines de réserve, effet à priori indirect car n'affectant pas une famille de protéines de réserve en particulier. Ces résultats suggérant un rôle négatif du gène *MtLIL* sur l'accumulation des protéines de réserve sont à l'inverse de ceux obtenus sur *LEC1*, ce qui nous permet d'émettre l'hypothèse d'un rôle antagoniste entre *MtLIL* et *MtLEC1* durant la phase de transition entre embryogenèse et remplissage.



Nos résultats ne nous permettent pas de conclure quant au rôle précis de *LIL* dans la régulation de la phase de remplissage, cependant l'analyse du profil d'expression du gène *MtLECI* nous permet d'émettre une hypothèse sur le rôle de ces deux gènes. En effet, il semble que les mutants *lec1* peuvent être complétés avec le gène *LIL* associé aux séquences régulatrices de *LECI* (Kwong *et al.* 2003 Yazawa *et al.* 2004; Alemanno *et al.* 2008). En revanche, ces mutants ne sont pas complétés par le gène *LIL* endogène, ce qui implique un rôle essentiel des régions régulatrices dans la fonction des gènes. Pour expliquer le fait qu'il n'y a pas complémentarité entre ces gènes, quatre hypothèses sont possibles : une intensité de *LIL* trop faible, une expression temporelle différente, une expression spatiale différente et enfin une régulation croisée entre *LIL* et *LECI* (Kwong *et al.* 2003). Dans notre étude, nous avons montré que ces deux gènes étaient exprimés de manière chevauchante et que *MtLIL* était 25 fois plus exprimé que *MtLECI*, ce qui exclut les deux premières hypothèses. La localisation d'expression a montré que ces gènes étaient exprimés dans les mêmes tissus mais n'a pas permis de déterminer une localisation plus fine au niveau cellulaire pour *MtLIL*, laissant la possibilité d'expression spatiale différente. En revanche l'hypothèse de régulation croisée entre les gènes est la plus probable. En effet, dans nos analyses d'expression du gène *MtLECI*, dans les plantes transformées sous-exprimant *MtLIL*, ont montré une dérégulation de ce gène, suggérant une régulation négative de *MtLIL* sur *MtLECI*. Cette régulation négative pourrait intervenir lors de la transition embryogénèse - remplissage et ainsi participer au déclenchement de la synthèse des protéines de réserve.

Ces hypothèses de travail restent encore à être validées en utilisant les mutants *lec1* identifiés dans la population TILLING. De plus, des doubles mutants *lec1/III*, en cours de réalisation, seront également un outil indispensable pour valider ces hypothèses.



# **MATÉRIELS ET MÉTHODES**



# CHAPITRE 6

## Matériel et Méthodes

### 1. MATERIEL BIOLOGIQUE

#### 1.1 Matériel végétal et conditions de culture

Les plantes de pois et de *M. truncatula* ont été mises en culture dans des pots contenant de la pouzzolane. Pour la production de graines matures, les plantes ont été cultivées en serre : température maintenue entre 14-30 °C pour le pois et 18-30°C pour *M. truncatula*, une photopériode de 16h. Une solution nutritive contenant 3,5:3,1:8,6 (N:P:K) est ajoutée cinq fois par semaine aux plantes tout au long de la culture.

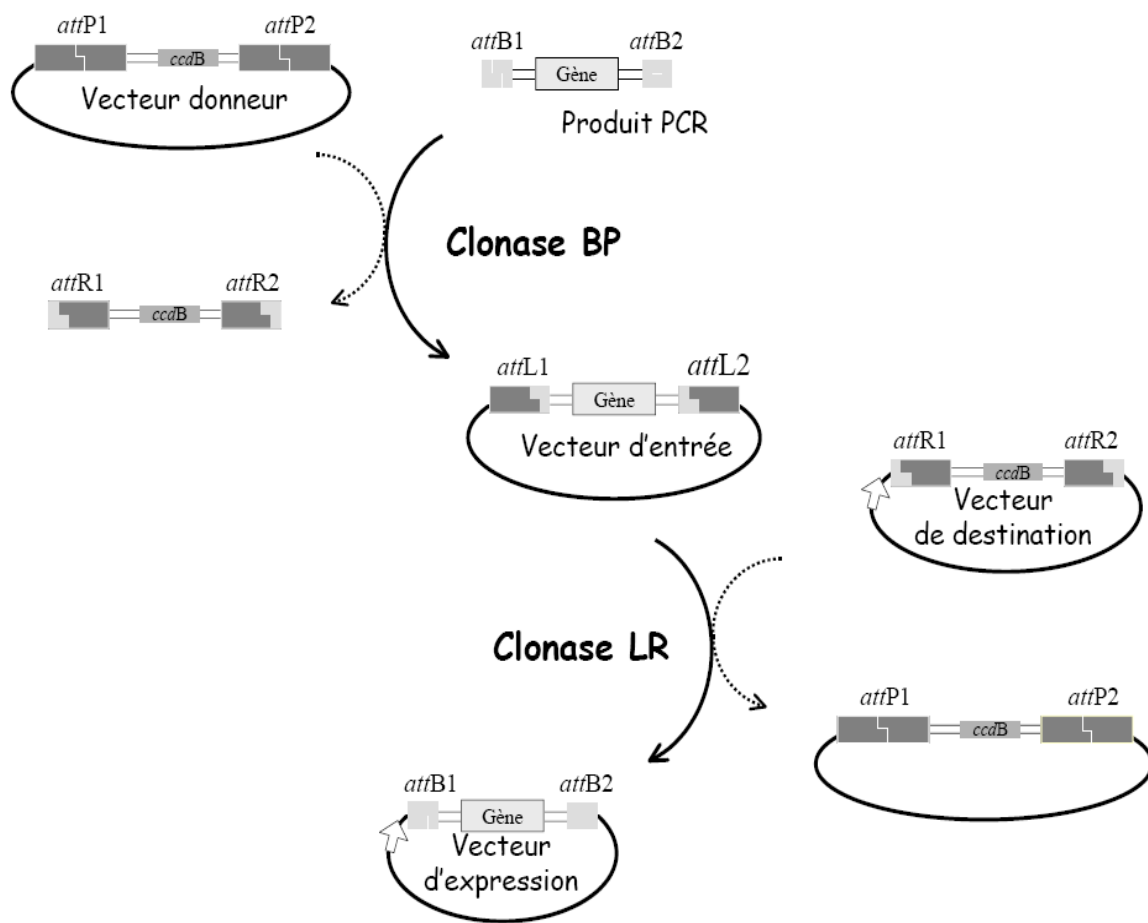
Les analyses de suivi d'expression sur le développement des graines ou les cinétiques de remplissage ont été réalisées à partir de plantes cultivées en conditions contrôlées (chambres climatisées). Ces plantes de *M. truncatula* (lignées A17 et R108 issues du cultivar Jemalong) ont été cultivées sous une photopériode de 16h, une intensité lumineuse de 220 $\mu$ E m<sup>2</sup> s<sup>-1</sup>, et une thermopériode de 22°C/19°C (Jour/Nuit). Les plantes de pois (cultivar 'Caméor') ont été cultivées sous une photopériode de 16h, une intensité lumineuse de 670  $\mu$ mol.m<sup>-2</sup> s<sup>-1</sup>, et une thermopériode de 24°C/16°C (Jour/Nuit).

#### 1.2. Souches bactériennes et milieux de culture

Les différentes souches de bactéries utilisées ainsi que leurs génotypes respectifs sont :

- *E. coli* souche DH5alpha (Invitrogen ref 18263-012) chimiquement compétente. Description du génotype : F-  $\phi$ 80*lacZ* $\Delta$ M15  $\Delta$ (*lacZYA-argF*)U169 *recA1 endA1 hsdR17*(rk-, mk+) *phoA supE44 thi-1 gyrA96 relA1  $\lambda$* -.
- *E. coli* souche DB3.1 (Invitrogen ref 11782-018) chimiquement compétente. Description du génotype : F<sup>-</sup> *gyrA462 endA1  $\Delta$* (sr1-*recA*) *mcrB mrr hsdS20*(r<sub>B</sub>-, m<sub>B</sub>-) *supE44 ara-14 galK2 lacY1 proA2 rpsL20*(Sm<sup>R</sup>) *xyl-5  $\lambda$* - *leu mtl1*.
- *E. coli* souche TOP10 (Invitrogen ref C404010) chimiquement compétente. Description du génotype : F- *mcrA  $\Delta$* (*mrr-hsdRMS-mcrBC*)  $\phi$ 80*lacZ* $\Delta$ M15  $\Delta$ *lacX74 recA1 araD139  $\Delta$* (*araleu*)7697 *galU galK rpsL* (Str<sup>R</sup>) *endA1 nupG*.





**Figure 6.1 : Principe du clonage GATEWAY (Invitrogen)** (Descriptif du mécanisme dans le texte section 1.3). Les régions *attB1*, *attB2*, *attP1*, *attP2*, *attR1* et *attR2* sont les régions de recombinaisons pour intégrer les inserts dans les vecteurs.

- La bactérie *Agrobacterium tumefaciens* C58pMP90, électro-compétente, est utilisée comme vecteur de transformation pour générer les plantes transgéniques. Elle possède un plasmide Ti ("Tumor inducing") désarmé contenant les gènes de virulence *Vir* permettant le transfert de l'ADN de Transfert (ADN-T) dans le noyau des cellules végétales.

Toutes les souches bactériennes sont cultivées dans du milieu LB (Bactotryptone 1% (p/v) ; extrait de levure 0,5% (p/v) ; NaCl 1% (p/v)) en présence ou non d'antibiotique et à la température appropriée (37°C pour *E. coli* ; 28°C pour *A. tumefaciens*).

Pour les milieux solides, 15 g.L<sup>-1</sup> d'agar bactériologique sont ajoutés. La stérilisation des milieux est réalisée par autoclavage pendant 20 min à 120°C.

### **1.3. Plasmides utilisés**

Les vecteurs de clonage, les plasmides recombinants ainsi que les constructions réalisées au cours de ce travail sont :

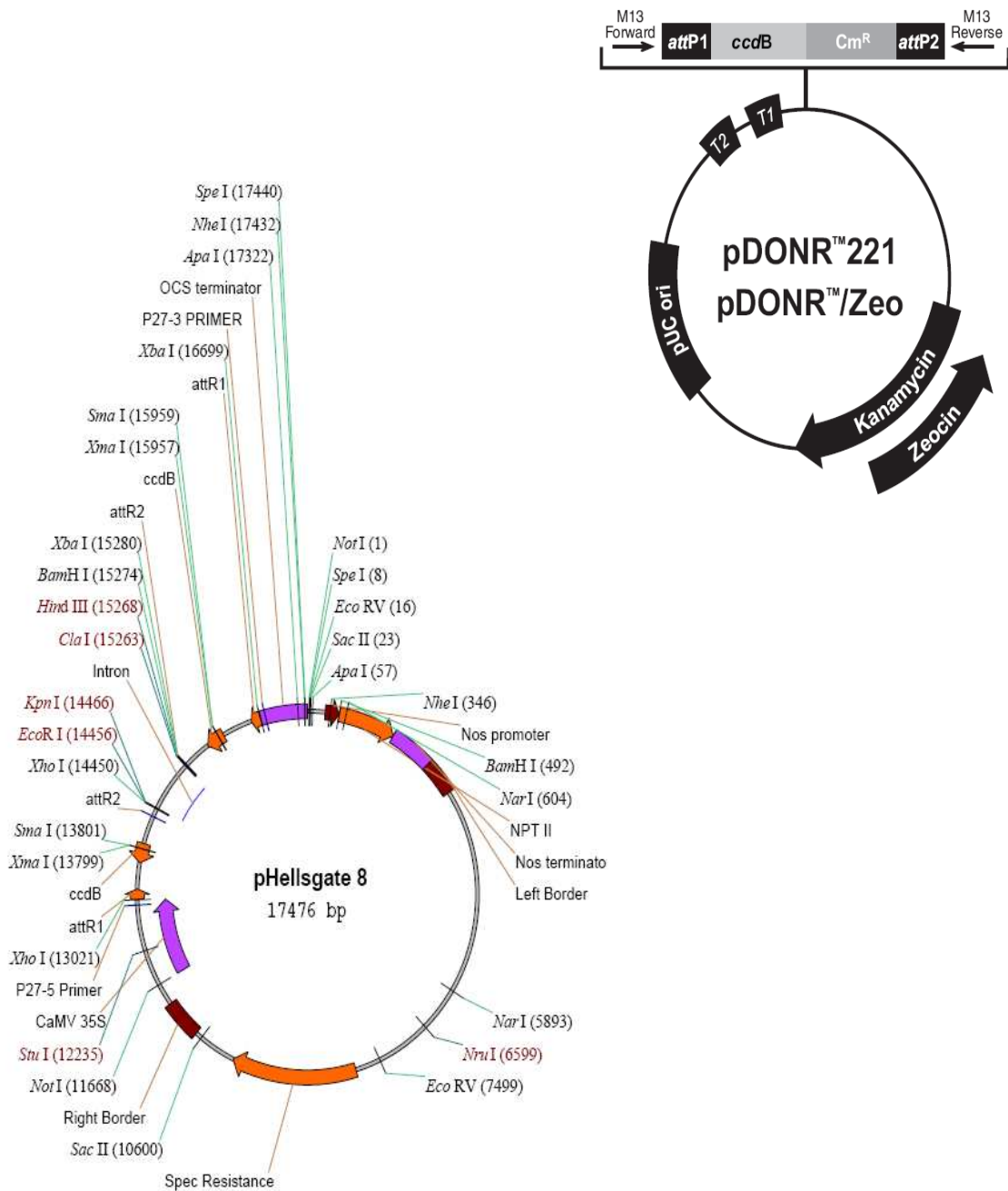
- Vecteur donneur pCR4 TOPO (Invitrogen ref. n°45-0030).

Il contient les gènes qui confèrent une résistance à l'ampicilline et à la kanamycine ; l'origine de réplication pUC ori pour permettre la réplication bactérienne ; le gène *ccdB* et le gène *LacZ* pour permettre un criblage des bactéries transformées. Le clonage des gènes dans ces plasmides est utilisé soit à des fins de séquençage ou soit pour la synthèse de sondes. Le clonage de ce plasmide est réalisé grâce au système Topo isomérase (Invitrogen).

- *Vecteurs d'expression utilisés pour la transformation des plantes :*

Pour générer les vecteurs d'expression, nous avons utilisé le système Gateway<sup>TM</sup>. Il est basé sur des réactions de recombinaison permettant l'intégration et l'excision du phage lambda dans le génome de *E. coli*. Cette méthode permet la recombinaison directionnelle d'un produit PCR (portant le gène d'intérêt et deux sites de recombinaison attB ajoutés grâce aux amorces de PCR) et d'un vecteur donneur (portant deux sites "attP") grâce à une réaction de recombinaison dite "BP". Le produit de cette première recombinaison, appelé vecteur d'entrée, sert de matrice à une seconde recombinaison (dite "LR") avec un vecteur de destination (portant deux sites de recombinaison "attR") pour obtenir un vecteur d'expression (Figure 6.1).

- Vecteur donneur pDONR 221 (4762bp), séquence référencée chez Invitrogen Ref. 12536-017. Ce vecteur est à la fois un vecteur donneur lors d'un clonage utilisant le système GATEWAY mais il a été également utilisé pour cloner des gènes à des fins de séquençage (Figure 6.2).



**Figure 6.2 :** Carte du vecteur donneur pDONR 221 et du vecteur destination pHELLSGATE 8.

- Vecteur destination pHELLSGATE 8 / Gateway Technology (Invitrogen). Ce vecteur a servi pour la transformation des plantes de *M. truncatula* R108 (transformations RNAi). Il contient la construction en "épingle à cheveu" sous le contrôle du promoteur 35S (Figure 6.2).

## 2. METHODES D'ANALYSE DES ACIDES NUCLEIQUES

### 2.1. Extraction des acides nucléiques

#### 2.1.1 Extraction d'ADN génomique

L'ADN génomique de plante est extrait selon la méthode de Dellaporta *et al.* (1983). Le matériel végétal (feuilles) est broyé à l'aide d'un pilon dans un mortier en présence d'azote liquide. Le broyat (5 g) est placé en présence de 10 mL de tampon d'extraction (Tris-HCl 100 mM pH 8, EDTA 50 mM pH 8, NaCl 500 mM,  $\beta$ -mercaptoéthanol 10 mM). Après addition de 2,7 mL de SDS 10% (p/v), le mélange est vortexé puis incubé pendant 10 min à 65°C afin de rompre les parois végétales. L'addition de 3,3 mL d'acétate de potassium 5 M pH 7, suivie d'une incubation de 20 min à 4°C, permet d'éliminer les protéines et les polysaccharides qui forment un précipité insoluble. Après centrifugation (20 min, 25000 g), le surnageant est récupéré et les acides nucléiques sont précipités à l'isopropanol pendant 30 min dans la glace. Après centrifugation (15 min, 20000 g), le culot d'acides nucléiques est séché puis repris dans 400  $\mu$ l d'une solution TE (Tris-HCl 50 mM pH 8, EDTA 10 mM pH 8). Le culot ainsi repris est soumis à une déprotéinisation par extraction au phénol-chloroforme (1:1) suivie d'une extraction au chloroforme-alcool isoamylique (24:1). Les acides nucléiques sont ensuite précipités à l'éthanol 95% (2,5 volumes) à -20°C, en présence d'acétate de sodium 0,3 M pH 5,8 (0,1 volume). Le culot final est rincé à l'éthanol 70%, séché sous vide et repris dans 400  $\mu$ l de TE. Un traitement à la RNase A (20  $\mu$ g/mL) est effectué, suivi d'une nouvelle déprotéinisation et précipitation à l'éthanol.

La concentration et la qualité de l'ADN génomique sont déterminées par spectrophotométrie à 260 nm et 280 nm, et après analyse par électrophorèse en gel d'agarose.

Extraction sur une quantité limitée de matériel végétal : l'ADN génomique est extrait selon la méthode modifiée d'Edwards *et al.* (1991). Une feuille est broyée quelques secondes avec un micro-pilon dans un tube Eppendorf. Puis, 200  $\mu$ l de tampon d'extraction sont ajoutés (Tris-HCl 0,2 M pH 7,5, NaCl 0,25 M, EDTA 25 mM, SDS 0,5% (p/v) et bisulfite de sodium 1%(p/v)) et homogénéisés par broyage pour quelques secondes supplémentaires. L'échantillon est centrifugé 5 min à 14000 g à température ambiante. Le surnageant est prélevé, mélangé à un volume d'isopropanol dans un nouveau tube et vortexé 5 secondes. Le mélange est laissé 5 min à température ambiante puis centrifugé 5 min à 14000 g. Le culot d'ADN est lavé avec 400  $\mu$ l d'éthanol 70%, séché 10 min à température ambiante puis repris dans 25  $\mu$ l de TE. Suite à une dernière centrifugation de 10 min à 14 000 g, le surnageant contenant l'ADN génomique est transféré dans un nouveau tube.



### 2.1.2 Extraction d'ADN plasmidique

L'ADN plasmidique est extrait en utilisant le kit " Genelute plasmid mini-prep " (Sigma), selon les recommandations du fournisseur. La culture bactérienne (5 mL) est sédimentée et le culot de cellules est soumis à l'extraction. L'ADN plasmidique est élué dans 50 µl d'eau ou de TE.

### 2.1.3 Extraction d'ARN totaux

Pour l'extraction et la manipulation des ARN, toutes les solutions sont traitées avec 0,1% (v/v) de DEPC, un inhibiteur des ribonucléases, pendant une nuit sous agitation ou préparées avec de l'eau traitée au DEPC. Tout le matériel est nettoyé au détergent Décon®, rincé, séché à l'éthanol et autoclavé.

Les broyages des différents échantillons ont été réalisés dans de l'azote liquide. Les ARN totaux ont été extraits selon la méthode modifiée de Chang *et al.* (1993) à partir de 0,2 à 1 g de poudre de matériel végétal.

On ajoute au matériel végétal congelé 15 ml de tampon d'extraction chauffé préalablement à 65°C. Ce tampon est composé de 2% de CTAB (Cétyltriméthylmethylammonium bromide), 0,1 M de Tris HCl, 0,5 M d'EDTA (Ethylene Diamine Tetra Acetic Acid, chélateur d'ions bivalents), 2 M de NaCl (créant une forte force ionique) et 2% de PVP (polyvinylpyrrolidone) servant à éliminer les polyphénols. Le principe de cette extraction est de séparer les protéines des acides nucléiques puisque le CTAB permet de précipiter les protéines et les polysaccharides tout en laissant les acides nucléiques solubles. Après agitation, on rajoute 2% de 2 β-mercaptoéthanol et on vortexe pendant 2 min. On incube la suspension à 65°C pendant 10 min. Ensuite, un volume de chloroforme est ajouté permettant la déprotéinisation de l'échantillon. Suite à une centrifugation à 6000g à température ambiante pendant 20 min, la phase supérieure est transférée dans un autre tube, cette étape est renouvelée une deuxième fois en ajoutant à nouveau un volume de chloroforme. La phase supérieure est transférée dans un autre tube, à laquelle 2,5 M de LiCl sont ajoutés. Les tubes sont mis à précipiter pendant toute la nuit à 4°C. Le LiCl permet la séparation entre l'ADN et l'ARN grâce à un gradient de densité dans lequel ne passeront que les ARN, les ADN ayant une densité plus faible ne traverseront pas ce gradient. Le lendemain, notre échantillon est centrifugé à 10000g, à 4°C pendant 30 min et le culot est repris avec 600 µL SSTE (SDS 0.5M, Tris-HCl 1M, EDTA 0.5M, NaCl 4M) chauffé préalablement à 65°C. On ajoute un volume de CHISAM (chloroforme-alcool isoamyl 24 :1) et après avoir vortexé pendant 30 secondes, les tubes sont centrifugés à 14000g à température ambiante durant 15 min. La phase supérieure est transférée dans un autre tube. Cette étape est renouvelée par ajout d'un volume de CHISAM. La phase supérieure est transférée dans un nouveau tube auquel on ajoute trois volume d'éthanol froid (-20°C) à 95% et on laisse précipiter l'ARN à -80°C pendant ¾ d'heure. Les tubes sont centrifugés à 14000g à 4°C pendant 30 min puis le culot est lavé à l'éthanol froid à 70% et les tubes sont recentrifugés à 14000g pendant 10 min. Enfin, le culot est séché à l'air et les ARN resuspendus avec 100 µL d'eau traité au DEPC.



Afin de déterminer la quantité d'ARN obtenue lors de l'extraction, des dosages sont réalisés en utilisant un spectrophotomètre (GeneQuant Pro, Amersham Pharmacia), en utilisant une dilution des échantillons au 1/100 dans du TE (Tris EDTA pH 8) sachant qu'une unité de DO correspond à 40 µg/ml d'ARN à 260 nm. Le spectrophotomètre calcule en plus de la concentration de nos ARNs, deux ratios indiquant le degré de pureté des : le premier ratio de **260 nm/230 nm** permet de déterminer la contamination de l'échantillon par des phénols, polyphénols et carbohydrates, ce ratio doit être supérieur à 2 sachant que 260 nm est la longueur d'onde d'absorbance des acides nucléiques et que 230 nm la longueur d'onde d'absorbance des composés contaminants. Le deuxième ratio **260 nm/280 nm** permet de déterminer la contamination en protéines de nos échantillons et doit être supérieur à 1,8 sachant que 280 nm est la longueur d'onde d'absorbance des cycles aromatiques des protéines.

## **2.2 Electrophorèse des acides nucléiques**

### **2.2.1 Electrophorèse d'ADN**

#### 2.2.1.1 Electrophorèse analytique

L'ADN restreint ou natif est analysé par électrophorèse horizontale non dénaturante en gel d'agarose [0,8 à 2% (p/v)] dans du tampon TAE 1X (Tris-HCl 40 mM pH 8, EDTA 1 mM, acétate de sodium 5 mM). Les échantillons sont additionnés de 0,2 volume de tampon de charge [glycérol 30% (v/v), bleu de bromophénol 0,25% (p/v), xylène cyanol 0,25% (p/v)]. Les fragments d'ADN sont visualisés sous lumière UV (BET; 0,5 µg.mL<sup>-1</sup>) après coloration du gel au bromure d'éthidium.

#### 2.2.1.2 Electrophorèse préparative.

Les électrophorèses préparatives sont réalisées en gel d'agarose 1% (p/v). Les fragments d'ADN sont découpés au scalpel après migration et purifiés en utilisant le kit "Ultrafree-DNA extraction from Agarose" de Millipore, selon les recommandations du fournisseur.

### **2.2.2 Electrophorèse d'ARN.**

Electrophorèses dénaturantes :

Les ARN sont analysés par électrophorèse horizontale dénaturante en gel d'agarose à 1,5% (p/v). L'agarose est amené à ébullition dans l'eau, puis refroidi à 50°C et complété avec une solution FRB 1X (MOPS 20 mM pH 7, acétate de sodium 8 mM, EDTA 1 mM) et 6% de formaldéhyde. Avant d'être déposés sur le gel, les échantillons sont dénaturés pendant 10 min à 65°C en présence de 10% (v/v) de FRB 5X, 50% (v/v) de formamide et 6,5% (v/v) de formaldéhyde. Les échantillons sont maintenus dans la glace et 0,2 volume de tampon de charge [glycérol 30% (v/v), bleu de bromophénol 0,25% (p/v), xylène cyanol 0,25% (p/v)] est ajouté. L'électrophorèse





est effectuée dans du tampon FRB 1X. Le gel est ensuite lavé pendant 30 min dans de l'eau avant d'être coloré dans une solution de BET à 0,5 µg/mL pendant 15 min. Les ARN sont visualisés sous lumière UV.

Electrophorèse non dénaturantes :

Les ARN sont analysés par électrophorèse horizontale en gel d'agarose à 1,5% (p/v). L'estimation de la qualité des ARN obtenus est observée suite à une migration sur gel d'agarose 1%. Les échantillons sont mélangés à du tampon de charge dénaturant FRB 1X contenant du MOPS, du formamide et du formaldéhyde, puis chauffés durant 5 min à 70°C pour favoriser la dénaturation des structures secondaires de l'ARN. La migration est réalisée dans du TAE 1X puis le gel est coloré dans une solution de BET à 0,5 µg/mL pendant 15 min. Les ARN sont visualisés sous lumière UV. Deux bandes doivent être clairement visibles, correspondant à l'ARN ribosomique 28S et 18S. Des échantillons de bonne qualité qui ne sont pas dégradés possèdent environ deux fois plus d'ARNr 28S que de 18S.

### **2.3 Clonage moléculaire**

Les fragments ADN sont clonés en utilisant le kit "TOPO TA cloning" en suivant les recommandations du fournisseur (Invitrogen).

La transformation des bactéries *E coli* chimiquement compétentes se fait par choc thermique en suivant les instructions du fournisseur (Invitrogen). Les bactéries sont mélangées avec le vecteur TOPO, dans lequel le Produit PCR d'intérêt a été préalablement inséré, pendant 30 min à 4°C, puis elles subissent un choc thermique de quelques dizaine de secondes à 42°C puis sont remises dans la glace. Ensuite, les bactéries sont immédiatement transférées dans du milieu SOC et soumises à une agitation pendant 1 h à 37°C. Une fraction de la culture est étalée sur un milieu LB supplémenté en antibiotique approprié, et, suivant le vecteur utilisé, le milieu LB est sélectif par la présence d'un inducteur de la β-galactosidase, l'IPTG (200 mg.mL<sup>-1</sup>) et d'un substrat coloré, le X-Gal [5% (p/v)].

### **2.4 Transcription inverse**

Les ARN totaux sont traités au préalable à la DNase RQI (Promega) pendant 30 min à 37°C suivant les instructions du fournisseur. Les échantillons d'ARN traités à la DNase sont concentrés sur Microcon YM-30 si besoin puis dosés au spectrophotomètre (voir section 2.1.3 Extraction ARN totaux). Pour vérifier l'efficacité du traitement, une PCR (4 min à 95°C et 39 cycles de 45 sec à 92°C, 45 sec au T<sub>m</sub> des amorces, 5 min à 72°C) est effectuée pour vérifier l'absence de contamination d'ADN génomique dans nos échantillons.



Pour la transcription inverse avec la SuperScript III (Invitrogen) :

Les ARN sont dénaturés à 65°C pendant 5 min puis la réaction de transcription inverse est effectuée à 50°C pendant 60 min dans le tampon " First Strand buffer 5X, Invitrogen ", en présence de 5ng/µl de random primers, 50µM d'oligonucléotides (oligo-dT<sub>20</sub>), de 500 µM de chaque dNTP, de 40 U de RNasin (Promega) et de 200 U de SuperScript™ III RNase H<sup>-</sup> Reverse Transcriptase (Invitrogen). L'enzyme est ensuite dénaturée pendant 10 min à 70°C.

Pour transcription inverse avec le kit iScript cDNA synthesis (Biorad) :

Pour 1 µg d'ARN total, le mélange réactionnel est composé de 4 µl de Tp 5X contenant des dNTP, des sels et des ions nécessaires pour l'activité de l'enzyme, des amorces polydT et des Random primers (amorces de quelques bases aléatoires), 1 µl d'enzyme et on complète à 15 µl contenant 2µg d'ARN dilué dans de l'eau DEPC. La synthèse des ADNc se fait dans les conditions suivantes : 5 min à 25°C pour l'accrochage des Random primers, 3 h à 42°C pour l'accrochage des amorces polydT et l'action de la Reverse transcriptase et enfin 5 min à 85°C pour inactiver l'enzyme.

Les fractions de transcription inverse sont ensuite utilisées en PCR pour amplifier le fragment souhaité.

## **2.5 Amplification génique ou PCR (Polymerase Chain Reaction)**

La technique de PCR décrite par Saiki *et al.* (1985) est composée de cycles d'amplification comprenant trois étapes : la dénaturation thermique de l'ADN bicaténaire matriciel, l'hybridation des amorces sur l'ADN matriciel, et la synthèse à partir des amorces, d'ADN complémentaire grâce à une ADN polymérase thermostable.

La réaction de PCR est réalisée soit après extraction de l'ADN, soit directement sur des ADNc ou des colonies bactériennes. Pour les PCR effectuées à partir d'ADN plasmidique, génomique ou d'ADNc, entre 8 et 100 ng d'ADN sont utilisés. Pour les PCR réalisées à partir de bactéries, une colonie est utilisée.

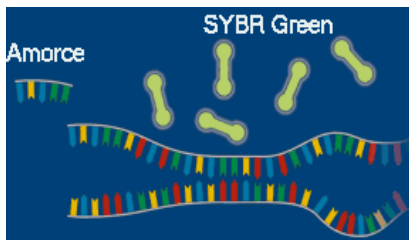
La composition de l'essai pour un mélange réactionnel de 10 à 100 µl est la suivante :

Tampon d'incubation 1X (Promega) (Tris-HCl 100 mM pH 8.3, MgCl<sub>2</sub> 15 mM, KCl 500 mM ; gélatine 0,01% (p/v)) ; 200 µM de chaque dNTP (dATP, dCTP, dGTP, dTTP) ; 0,2 µM d'amorces oligonucléotides "forward" et "reverse" et de 0,5 à 2U de Taq ADN polymérase (Q-BIOgene DNA polymérase)

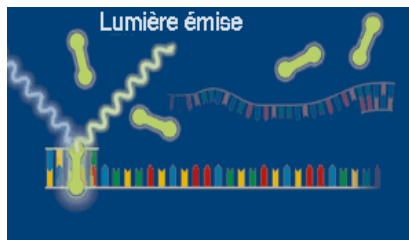
Les essais sont réalisés dans des tubes Eppendorf de 200 ou 500 µl ou sur des plaques de 96 ou 384 puits.

Conditions de PCR : les amplifications en routine sont réalisées dans un appareil automatisé (MJ Research PTC-200). Le mélange est soumis à des cycles d'amplification (30 à 40 cycles) qui comprennent une étape de dénaturation des matrices de 30-60 s à 95°C, une étape d'hybridation dont la durée et la température varient

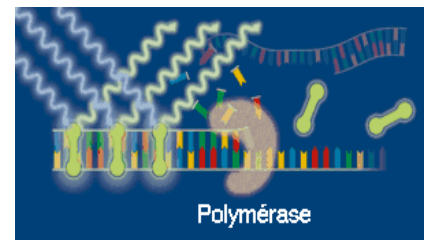
(a)



1- Au début de l'amplification, le mélange réactionnel contient de l'ADN dénaturé, les amorces et le fluorophore non lié.

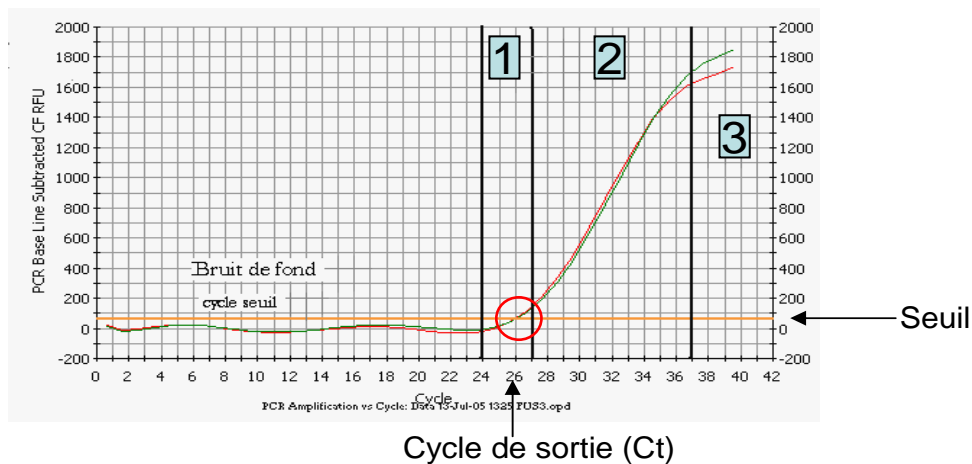


2- Après le couplage des amorces, le fluorophore se lie au double brin. La liaison à l'ADN se traduit par une augmentation de la fluorescence.



3- Pendant l'étape d'élongation, le nombre de fluorophores liés à l'ADN synthétisé augmente ce qui se traduit par une augmentation de la fluorescence.

(b)



**Figure 6.3 : (a) Principe de la quantification en temps réel**

(<http://www.anapath.necker.fr/metanapat/PCRWWW/PCR>) et **exemple de courbe d'amplification de 2 échantillons**. Les étapes d'amplification exponentielle, d'amplification linéaire et de saturation sont indiquées, ainsi que le seuil et le cycle de sortie (threshold cycle, Ct) sont représentés pour ces échantillons.

selon les amorces utilisées, et une étape de synthèse d'ADN par extension d'amorces réalisée à 72°C et dont la durée varie avec la taille de la molécule d'ADN à amplifier.

### **2.6 RT-PCR Quantitative**

Principe de la méthode :

La PCR quantitative (qPCR) ou PCR quantitative fluorescente en temps réel est basée sur le principe de la PCR classique, mais contrairement à celle-ci, la qPCR nous permet de suivre en temps réel au cours des cycles d'amplification la fluorescence, proportionnelle au nombre de molécules synthétisées grâce à la détection et à la quantification d'un agent intercalant fluorescent, le SYBR Green. Le SYBR Green est un fluorophore capable de se lier à l'ADN double brin. Ce fluorophore n'est pas fluorescent en solution sous forme libre, mais fluoresce lorsqu'il est lié à l'ADN (Figure 6.3). Lorsque l'on suit en temps réel l'amplification d'une séquence spécifique, l'émission de fluorescence est mesurée à la fin de chaque étape d'élongation pour chacun des cycles d'amplification. Cette mesure est réalisée par un système optique de lecture intégré à l'appareil de qPCR permettant de suivre l'augmentation en temps réel de la quantité d'ADN amplifiée durant la réaction lors de la phase d'amplification exponentielle, phase où l'amplification suit la formule : quantité =  $2^n$  (n correspondant au nombre de cycles) (Figure 6.3). Le seuil de détection correspond à une intensité de fluorescence significativement plus élevée que ce niveau de base. Le cycle seuil ou threshold cycle (Ct) correspond au nombre de cycle PCR nécessaire pour que la fluorescence soit égale au seuil de détection. Le Ct est inversement proportionnel à la quantité d'ADN cible présent dans l'échantillon avant la PCR.

La méthode utilisée de quantification relative nécessite l'utilisation d'un ou de plusieurs gènes de référence étudiés dans les mêmes conditions que le gène cible. Il doit présenter une expression stable au cours du processus étudié ou dans les tissus étudiés. Le choix du gène de référence doit être judicieusement choisi car la quantification relative dépend directement de l'expression du gène de référence. Dans cette thèse, plusieurs gènes contrôles ont été utilisés pour normaliser les données (cf matériel et méthodes des publications présentes dans le manuscrit). Chez *M. truncatula*, nous avons utilisé *MSC27* (TC85211) pour le développement de la graine et pour l'étude dans les trois tissus de la graine ainsi que la moyenne géométrique de quatre gènes pour l'expérience PCR quantitative à haut débit : *PDF2* (TC107161), *GAPDH* (MtC00030\_GC), *PTB* (TC111751), *ACTIN* (TC107326). Chez *P. sativum*, nous avons utilisé la moyenne géométrique de deux gènes *HISTONE H1* (#X05636) et *ELONGATION FACTOR1 BETA* (X96555) pour la normalisation des expression dans la graine.

Les réactions de PCR quantitative ont été réalisées réalisés dans un thermocycleur iCycler iQ (Biorad) ou ABI7900HT (Applied Biosystems) ou LC480 (Roche Applied science). Trois "master mix" de qPCR ont été utilisés : iQ SYBR Green Supermix (Biorad), 2X SYBR Green mix reagent (Applied Biosystems) et le MESA Green supermix (Eurogentec) en suivant les recommandations des fournisseurs.

Les quantités relatives d'ADNc ont été mesurées selon la méthode du  $\Delta\Delta C_t$  ou selon la méthode de courbe



standard. Les quantités relatives ont été normalisées par rapport au(x) gène(s) de référence. Les réactions ont été réalisées en duplicat ou triplicat et la qualité de l'amplification a été vérifiée après les cycles de PCR, grâce à une courbe de fusion.

## **2.7 Hybridation in situ**

L'hybridation *in situ* a été réalisée sur des coupes de graines à 14 JAP.

Les sondes ARN sens et antisens marquées à la digoxygénine ont été synthétisées grâce au kit Riboprobe Combination system T3/T7 (Promega). Le fragment cloné qui a servi de matrice pour la synthèse des sondes ARN sens (contrôle de l'expérimentation) et antisens correspondait au domaine C du gène *MtLIL*. Les fragments linéarisés sont purifiés après électrophorèse préparative. Le marquage des sondes est réalisé dans un mélange réactionnel de 20 µl composé de 1 µg d'ADN plasmidique linéarisé, de Tris-HCl 40 mM pH 8, de MgCl<sub>2</sub> 6 mM, de DTT 10 mM, de spermidine 2 mM, d'ATP, CTP, GTP (1 mM chacun), d'UTP 0,5 mM, de Dig-11-UTP 0,5 mM et de 40 U des ARN polymérase correspondantes (T3 ou T7). L'incubation est effectuée dans un thermocycleur durant 1 h à 37°C. La réaction est stoppée par addition de 5 µl d'EDTA 0,2 M pH 8. Les ARN synthétisés sont précipités en présence de d'acétate d'ammonium et de 3 volumes d'éthanol absolu pendant 45 min à -80°C, lavés à l'éthanol 70%, séchés à température ambiante puis repris dans 20 µl d'eau DEPC. Les concentrations d'ARN synthétisés sont déterminées par comparaison à une gamme étalon d'ARN témoin.

Fixation et traitement du matériel biologique :

Les solutions de fixation et celles nécessaires au traitement des coupes (hormis les solutions d'alcool, d'Histosol Plus (CML) et de paraffine (ParaplastPlus, SIGMA)) sont préparées extemporanément.

Le matériel végétal est plongé dans une solution de formaldéhyde. La fixation est réalisée à température ambiante pendant 15 min sous vide puis sur la nuit à 4°C. Les étapes de déshydratation sont réalisées dans des bains d'alcool de concentration croissante de 10, 30, 50, 70, 95 % pendant 1h pour chaque bain puis la nuit à 100%. Puis, le matériel végétal doit subir une imprégnation par un solvant de la paraffine, l'histoclear. Cette imprégnation doit être progressive : 2:1 (2vol d'éthanol/1vol histoclear) pendant 2h, 1:1 pendant 2h, 1:2 pendant 2h puis Histoclear seul pendant 1h. Enfin le matériel doit être inclus dans la paraffine chauffée à 60°C en changeant trois fois les bains de paraffine toutes les 3h. Les préparations sont ensuite séchées dans une étuve à 45°C pendant 24 h.

Déparaffinage et réhydratation des coupes :

Les lames sont tout d'abord déparaffinées par 3 bains successifs d'Histoclear pendant 10 min, puis rincées dans 2 bains d'alcool 100% pendant 10 min. Les opérations de réhydratation sont réalisées dans des bains d'alcool de 95, 70, 50 et 30% pendant 1 min chacun. Les lames sont ensuite placées dans 3 bains successifs de PBS 1X pendant 5 min, et enfin dans un bain de SSPE 2X pendant 10 min.





Préhybridation et hybridation :

Une solution de préhybridation (formamide 50% (v/v), SSPE 4X, Dextran sulfate 10% (p/v), Denhardt 1X, ARNt 250  $\mu\text{g.mL}^{-1}$ , ADN de sperme de saumon soniqué 250  $\mu\text{g.mL}^{-1}$ ) est déposée sur les coupes à raison de 150  $\mu\text{l}$  par lame. La préhybridation se déroule pendant 1 h, dans une chambre de préhybridation DAKO déposée sur la lame. Les préparations sont ensuite rincées dans un bain de SSPE 2X. 100  $\mu\text{l}$  de la solution de préhybridation contenant 50 ng de sonde marquée et dénaturée 10 min à 80°C sont déposés sur chaque lame. Les lames sont recouvertes d'une lamelle siliconée puis placées dans un incubateur à 45°C pendant une nuit.

Lavages et détection immunologique :

Après hybridation, l'excès de sonde et les hybrides aspécifiques sont éliminés par des lavages de 30 min à température ambiante dans du SSPE à concentration saline décroissante (2X puis 1X et 0,5X), puis les préparations subissent un dernier lavage à 55°C dans du SSPE 0,5X. Suite à ce lavage stringent, les lames sont soumises à un dernier lavage en SSPE 0,5X pendant 30 min, à température ambiante. L'étape de détection immunologique à l'aide d'anticorps anti-digoxygénine (Roche Biochemical) nécessite des traitements préalables des lames. Elles sont immergées 2 fois dans du tampon 1 (Tris-HCl 0,1 M, NaCl 0,15 M) à température ambiante pendant 5 min, et 1 fois 30 min dans une solution de BSA [Triton X-100 0,3% (v/v), BSA 2% (p/v)] préparée dans du tampon 1, à température ambiante. 150  $\mu\text{l}$  de la solution d'anticorps diluée au 1/500 dans du tampon BSA sont déposés sur chaque lame. L'incubation dure une nuit à 4°C. L'excès d'anticorps est éliminé dans deux bains successifs de 15 min contenant 200 mL de tampon 1 suivis de 2 rinçages de 5 min en tampon 3 (Tris-HCl 100 mM pH 9,5, NaCl 100 mM,  $\text{MgCl}_2$  50 mM), à température ambiante sous agitation douce. 150  $\mu\text{l}$  de la solution de NBT/BCIP (Roche Biochemical) (respectivement 450  $\mu\text{g.mL}^{-1}$  et 175  $\mu\text{g.mL}^{-1}$  dilués dans du tampon 3) sont déposés sur les coupes qui sont alors placées à l'obscurité. La réaction de coloration, se traduisant par l'apparition d'une coloration rouge-violette est arrêtée par rinçage dans un bain d'eau distillée. Les lames sont mises à sécher une nuit à température ambiante et les préparations sont recouvertes d'une lamelle pour l'observation au microscope optique.

## **2.8 Séquençage des acides nucléiques**

Le séquençage des fragments clonés ou obtenus par PCR ont été réalisés par la société MWG Sequencing Services ou par la société Millegen.

## **2.9 RACE-PCR (Rapid Amplification of 5' cDNA Ends)**

L'identification des extrémités 5' des fragments d'ADNc a été réalisée en utilisant le kit "5' RACE System" (Invitrogen) en suivant les recommandations du fournisseur.



### 3. METHODES D'ANALYSE DES PROTEINES

#### **3.1 Extraction des protéines totales**

Les protéines totales ont été extraites à partir de graines matures avec 500µl de tampon de lyse pour 10mg de poudre de graines matures (urée 7,4M ; thiourée 2,1M ; CHAPS 62mM ; DTT 8mM ; pharmalytes 1% ; Trizma hydrochloride 21mM ; Trizma base 16,5M ; Triton X-100 0.24% ; inhibiteurs de protéases 1X ; DNase I 60U ; RNaseA 72µg/ml) selon un protocole établi pour les graines *d'A. thaliana* (Gallardo et al., 2002). Cet homogénat est agité pendant 40 minutes à température ambiante, puis centrifugé 10 min à 4°C à 20000g. Le surnageant correspondant à l'extrait protéique total est soumis à une seconde centrifugation et stocké à -20°C.

#### **3.2 Dosage des protéines**

La concentration en protéines dans les différents extraits est déterminée selon la méthode colorimétrique de Bradford (1976). Le principe du dosage des protéines par la méthode Bradford repose sur la fixation non covalente d'un colorant sur les protéines, le bleu de Coomassie. A pH acide, la forme anionique de ce colorant se fixe préférentiellement aux protéines par des interactions électrostatiques avec les groupements cationiques des acides aminés, principalement l'arginine, mais aussi la lysine et l'histidine. Le complexe ainsi formé absorbe dans une gamme de longueurs d'onde dont le maximum se situe à 595 nm. Ce dosage a été réalisé à l'aide d'un réactif commercial (Bio-Rad Protein Assay), compatible avec des conditions dénaturantes et réductrices, et d'une gamme étalon (de 0 à 0.2µg/µl) de BSA (Bovin Serum Albumin). Pour minimiser toute interférence possible avec le test, les extraits protéiques sont dilués 200 fois dans l'eau milliQ et la gamme BSA est réalisée en présence du tampon utilisé lors de l'extraction à la même dilution que les essais. 1 ml de réactif Biorad dilué au 1/5 fut ajouté aux 100 µl de chaque échantillon à mesurer.

Pour les travaux de cette thèse, les dépôts de protéines sur gels 1D et 2D ont été faits à quantité constante de protéines.

#### **3.3 Electrophorèse monodimensionnelle**

Les échantillons sont dilués dans le tampon de charge [Tris-HCl 62 mM pH 6,8, glycérol 30% (v/v), SDS 2% (p/v), bleu de bromophénol 1% (p/v)] et dénaturés à 95°C pendant 5 min. Les échantillons sont ensuite déposés sur gels SDS-PAGE constitués : d'un gel de résolution : Tris-HCl 0.3M pH 8.8, Acrylamide-bisacrylamide 12% ou 14%, SDS 0.1%, Persulfate d'ammonium 0.05%, Temed 0.05% et d'un gel de concentration : Tris-HCl 0.125M pH 6.8, Acryl-bis 4%, SDS 0.1%, Persulfate d'ammonium 0.05%, Temed 0.1%. La migration est réalisée pendant 1h30 à Ampérage constant (30mA) dans le tampon d'électrophorèse suivant : Trizma base 25 mM, glycine 0.2M, SDS 0.1%.



Après avoir coloré le gel au bleu de Coomassie (BioRad Biosafe) et numérisé les gels avec un scanner Odyssey (LiCor), les bandes protéiques peuvent être quantifiées par densitométrie en utilisant le logiciel Quantity One (BioRad) ou Odyssey (LiCor).

### **3.4 Electrophorèse bidimensionnelle**

#### 3.4.1 Principe

L'électrophorèse bidimensionnelle (2-DE), décrite par O'Farrell (1975), est une méthode combinant deux séparations séquentielles en gel de polyacrylamide :

i) la première correspond à une séparation des protéines selon leur point isoélectrique (pI) (IEF=isoélectrofocalisation) dans un gel contenant un gradient de pH immobilisé réalisé par coulage de solutions d'Immobilines co-polymérisant avec le gel d'acrylamide. Les protéines soumises à un champ électrique constant migrent dans le gel jusqu'à la position où le pH du milieu est égal au pI de la protéine pour laquelle sa charge globale est nulle.

ii) la seconde dimension correspond à une séparation en SDS-PAGE. Entre les deux se situe une étape dite d'équilibration des gels d'isoélectrofocalisation qui permet aux polypeptides de s'alkyler et se charger en SDS (sodium dodécyl-sulfate). Le SDS interagit avec les protéines et donne à chacune un même rapport charge/masse. Dès lors, les molécules sont séparées en fonction de leur masse apparente (kDa). Les spots sont ensuite révélés au bleu de Coomassie.

#### 3.4.2 Isoélectrofocalisation (= 1<sup>ère</sup> dimension)

Les protéines sont donc dans un premier temps séparées en fonction de leur charge lors de l'isoélectrofocalisation (IEF), en présence d'urée (non ionique) et de détergents ne modifiant pas la charge des protéines. L'IEF est réalisée en gradient de pH immobilisé (IPG strips pI 3-10 non linéaire de 24 cm), à partir d'une quantité constante de protéines (200 µg) de chaque extrait protéique. Chaque échantillon est complété à 460µl avec une solution de solubilisation (urée 7M, thiourée 2M, Triton X-100 2% (v/v), CHAPS 58 mM, DTT 20mM, ampholytes (pH3 à 10) 1% (p/v)) contenant une trace de bleu de bromophénol pour vérifier la séparation électrophorétique des protéines au cours de l'IEF. Chaque homogénat est ensuite réparti uniformément dans la chambre de réhydratation des strips (Reswelling tray, Amersham Biosciences). Les strips pI 3-10 NL sont ensuite déposés au contact des extraits. Pour éviter toute évaporation, de l'huile minérale est ajoutée dans chaque rangée. Les strips sont ensuite réhydratés dans le système IPGphor pendant 7h à 0V (volts) suivi de 7h à 50V pour éliminer les sels au contact des extraits protéiques. La migration de l'IEF est réalisée à 22°C durant 1 h à 350V, 3 h à 3500V et 7 h à 8000V.

#### 3.4.3 Equilibration des strips

Après la séparation des protéines en première dimension par IEF, celles-ci sont équilibrées en charges négatives



en présence d'un détergent anionique, le sodium-dodecyl-sulfate (SDS), et d'un agent réducteur des ponts disulfures, le DTT. Ainsi leur mobilité électrophorétique lors de la seconde dimension dépendra uniquement de leur masse moléculaire. Une solution réductrice constituée de DTT 50mM, de SDS 2.5% (p/v), d'urée 6M, de bis-tris 0.15M, d'HCl 0.1M et de glycérol (30%, v/v) est mise au contact des strips pendant 15 minutes sous agitation douce. Les strips sont ensuite rincés avec la même solution dépourvue de DTT pendant 1 minute, puis une solution alkylante [iodoacétamide 220mM, SDS 2.5% (p/v), urée 6M, bis-tris 0.15M, HCl 0.1M et glycérol (30%, v/v)] est ajoutée pendant 15 minutes sous agitation douce. L'iodoacétamide est un agent d'alkylation des thiols libres qui stabilise les réductions réalisées par le DTT.

#### 3.4.4 Séparation sur gels SDS PAGE (=2<sup>ème</sup> dimension) et coloration

La seconde dimension est effectuée verticalement en gels SDS-PAGE [acrylamide 10% (v/v), pipérazine diacrylamide 0.33%, (p/v) (PDA), Trizma base 165mM, HCl 0.1M, persulfate d'ammonium 0.07% (p/v), Temed 0.035% (v/v)]. Après polymérisation, une solution d'agarose dénaturante [agarose " low melting " 1% (p/v), SDS 4%, bis-tris 0.15M, HCl 0.1M, trace de bleu de bromophénol (pour suivre le front de migration)] est ajoutée dans les puits. Les strips (gels de 1<sup>ère</sup> dimension), équilibrés en SDS, sont ensuite rapidement déposés dans l'agarose. Après solidification de l'agarose, un marqueur de taille (Low Range, Bio-Rad) est déposé sur les gels. L'électrophorèse est réalisée à 10°C, dans le système Hoefer Dalt (Amersham Biosciences), en présence d'un tampon (pH 8.0) contenant trizma base 25mM, taurine 200mM et SDS 0.1% (p/v), pendant 1h à 35V et 14h à 80V. Après migration, les gels sont rincés deux fois cinq minutes dans l'eau milliQ, colorés pendant 2 heures avec du bleu de Coomassie (Biosafe Coomassie, Coomassie Blue G250 Stain), qui présente une grande affinité pour les protéines en milieu acide. Cette propriété permet de combiner la coloration des protéines et leur fixation. Les gels sont ensuite lavés au minimum deux fois à l'eau milliQ pour éliminer le bruit de fond.

#### 3.4.5 Analyse des gels 2D

Pour déterminer les modifications du taux d'accumulation des protéines, les images des gels sont analysées de manière quantitative. Les gels analytiques sont numérisés avec le scanner ODYSSEY (LiCor). Les paramètres d'acquisition de l'image sont réglés par l'intermédiaire de ce logiciel. Les analyses quantitative sont réalisées en utilisant le logiciel Progenesis SameSpots (Non Linear Dynamics).

Alignement automatique des gels (" matching ") : les gels sont comparés entre eux en utilisant le mode "matching" qui détermine automatiquement les spots communs entre les différents gels. Un gel de bonne qualité est choisi comme référence. Tous les gels sont ensuite automatiquement comparés à cette référence. Les spots protéiques communs dans les différents gels sont reconnus de manière automatique. Une correction manuelle de l'alignement des spots est ensuite réalisée à l'aide des outils disponibles permettant de rajouter, d'enlever et de séparer des spots. Cette étape est la plus critique, l'alignement doit être parfait pour obtenir une bonne analyse quantitative.





Détection des spots protéiques : cette étape consiste à détecter les spots protéiques sur les gels 2DE. Une bonne détection représente l'ensemble des spots entourés avec précision. La détection des spots est réalisée dans un premier temps en mode automatique puis est ensuite vérifiée sur chaque gel.

Soustraction du bruit de fond : le bruit de fond des gels est soustrait du volume des spots. Le mode "background subtraction" permet de soustraire ce bruit de fond selon le mode "Lowest on Boundary" qui soustrait le bruit de fond le plus faible localisé à la périphérie de chaque spot.

Normalisation du volume des spots : les volumes des spots sont fonction de la densité optique et de l'aire des spots. Ces volumes sont normalisés en utilisant la méthode "Total Spot Volume" qui normalise le volume de chaque spot par rapport à l'aire totale représentée par l'ensemble des spots.

Gels moyens ("averaged gels") et expression des résultats : des gels moyens sont créés à partir des gels de chaque échantillon. Le logiciel calcule automatiquement les volumes normalisés moyens. Les valeurs de spots peuvent ensuite être directement exportées dans le logiciel Excel. Nous appliquons ensuite des tests statistiques (ANOVA à un facteur) pour identifier les spots qui varient selon un critère de p-value < 5% et une puissance de test statistique  $P > 75\%$  (Statistica, Statsoft ou R).

#### 3.4.6 Identification des spots protéiques

La spectrométrie de masse permet aujourd'hui, grâce à des progrès considérables en instrumentation et en méthodologie, d'identifier rapidement une protéine isolée par exemple par électrophorèse. La séquence de la protéine à analyser est identifiée grâce aux masses caractéristiques des peptides (empreintes peptidiques massiques, MALDI-TOF) ou suite à une fragmentation de peptides (Informations de séquence, MS-MS). L'identification d'une protéine est réalisée par comparaison avec les séquences contenues dans les bases de données publiques ou privées. C'est aujourd'hui un outil incontournable pour l'analyse des protéines (Wilm, 2000).

Pour éviter toute contamination, notamment par la kératine, le prélèvement des spots est réalisé dans un espace confiné. Chaque spot est excisé par l'intermédiaire de l'extrémité d'un cône stérile, déposé dans un tube Eppendorf autoclavé. Une fois prélevés, les spots protéiques sont conservés à  $-80^{\circ}\text{C}$  jusqu'à leur analyse par spectrométrie de masse.

Le séquençage des peptides a été réalisé par la plate-forme INRA de spectrométrie de masse de Nouzilly (V. Labas, INRA Tours) par la méthode LC 1D nanoESI MS/MS. Après l'hydrolyse enzymatique par la trypsine, l'échantillon est désalé puis injecté en LC-MS/MS en utilisant une nanoHPLC en ligne avec un spectromètre de masse de type ESI-Q-TOF. Chaque peptide est alors fragmenté dans le spectromètre de masse ; l'analyse des fragments permet de déterminer les éléments de la séquence en acides aminés du peptide, séquence qui sera utilisée pour identifier la protéine dans les banques de données. Le logiciel Bioworks permet d'intégrer les profils massiques et de déduire les séquences en acides aminés des peptides. Ces séquences peptidiques sont alors soumises aux bases de données protéiques à travers des logiciels de BLAST (Basic Local Alignment



Search Tool) sur NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>), MtGI v8 (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=medicago>) et sur des bases de données privées de pois afin d'identifier la protéine correspondante.

Protocole fourni par le laboratoire d'analyse :

Les échantillons ont été digérés dans le gel par la trypsine (Trypsin Sequencing Grade, Roche, référence 1 418 475) pendant 2h à 37°C. Les peptides issus de la digestion des spots sont analysés par un système nanoHPLC Ettan MDLC (GE HealthCare, Germany) piloté par le logiciel UNICORN™ couplé à un spectromètre de masse LTQ de type trappe à ion linéaire (Thermo Electron, US). Les échantillons sont dessalés et concentrés en ligne par une précolonne Zorbax 300-SB C18 trap (5µm, 5 X 0.3 mm) (Agilent Technologies, Germany) puis séparés sur une colonne analytique phase réverse Zorbax 300-SB C18 trap (3.5µm, 150 X 0.075 mm) (Agilent Technologies, Germany) avec un débit de 320 nL/min.

Le profil du gradient est le suivant :

- Equilibration des colonnes avec 95% de solvant A (0.1% acide formique) et 5% de solvant B (84% acétonitrile/15,9 % H<sub>2</sub>O /0.1% acide formique, v/v)
- Gradient de 5 à 15% B en 2 min
- Gradient de 15 à 50% B en 30 min,
- Palier à 95% de B pendant 10 min.

L'acquisition des données se fait de façon automatique entre les modes MS et MS/MS (fragmentation) : un " enhanced centroid full scan " (m/z 500-2000) est suivi par 3 MS/MS centroid scans sur les 3 pics détectés les plus intenses (Qz 0.25, activation time 30 ms, collision energy 40). L'exclusion dynamique est activée durant 30 sec avec un " repeat count " de 1.

Les spectres MS/MS sont confrontés à 3 banques de donnée à l'aide du logiciel Bioworks version 3.3.1 en considérant les critères suivants :

- Banque de donnée indexée : nr NCBI, TIGR8 et contig-annot\_Pois\_v4
- Trypsine
- 2 miss cleavages
- 1.4 Da de précision de masse sur MS et 1 Da sur MSMS
- carbamidométhylation et oxydation des méthionines en modifications variables
- filtre Xcorr vs Charge state 1.70 (1+), 2.0 (2+) 2.5 (3+) 3.0 (4+)
- 2 peptides différents au minimum
- Probabilité (peptides) < 0,5



### **3.5 FPLC (Fast Protein Liquid Chromatography)**

Les protéines sont extraites, à partir de 100 mg de poudre de graines matures, avec une solution d'extraction non dénaturante pH 8,5 (Tris 50mM ; K<sub>2</sub>SO<sub>4</sub> 5% ; EDTA 1mM ; -mercapto 0,05% ; benzamidine 1μM ; leupeptine 1μM ; PMSF 1μM) après 1h d'agitation et une centrifugation de 20 min à 6000g à 4°C. Le surnageant est filtré (0,22μm) et dilué au demi dans une solution de Tris 50mM pH8,5 pour être congelé à -80°C. Les séparations en FPLC sont réalisées avec un système Waters (600<sup>E</sup> system controller, 717 injector, 486 detector) suivant le protocole de Baniel *et al.* (1998) à partir de 75μl d'extrait protéique filtré. La colonne de séparation est une MonoQ 5/50 GL analytical column (Amersham Biosciences) et l'élution des protéines se fait grâce à un gradient d'élution de NaCl de 0 à 0,72M (40 min à 1ml/min). Les fractions protéiques sont détectées en utilisant l'absorbance des liaisons peptidiques à 214nm. La quantification de chaque fraction est réalisée par intégration des surfaces de pics.

## **4. METHODE DE TRANSGENESE ET D'ANALYSES DES PLANTES TRANSGENIQUES.**

### **4.1 Construction des vecteurs de transformation**

La séquence codante du gène *MtLIL* identifiée chez *M. truncatula* R108 correspondante à une partie du domaine B et au domaine C a été introduite en orientation sens ou antisens dans le plasmide pHELLSGATE8 sous le contrôle du promoteur 35S. Ce clonage a été réalisé grâce au système Gateway (Invitrogen) en passant tout d'abord dans le plasmide donneur pDONR puis par recombinaison homologue dans le plasmide destination, le pHELLSGATE8 (Figure 6.1 et 6.2).

Les constructions réalisées sont introduites dans la souche *Agrobacterium tumefaciens* C58pMP90 par la technique d'électroporation. Les souches transformées sont cultivées pendant 48 h à 28°C, sous agitation, dans le milieu sélectif LB additionné de spectinomycine. Cette suspension bactérienne est diluée au centième, remise en culture pendant une nuit avant d'être utilisée pour la transformation.

### **4.2 Transformation et sélection des plantes**

La transformation a été réalisée sur des explants de feuille de *M. truncatula* Jemalong lignée R108 par S. Djennane, L. Jacas et S. Ochatt (UMR-LEG).

Les explants de feuilles sont coupés afin d'induire une meilleure réponse à la blessure nécessaire pour l'infection par *A. tumefaciens*. Les explants sont immergés pendant une trentaine de minutes dans une culture en phase exponentielle de croissance d'*A. tumefaciens* transformées par les plasmides recombinants contenant les constructions. L'excès de cellules bactériennes est éliminé entre deux feuilles de papier absorbant stérile. Le matériel végétal est ensuite co-cultivé avec les agrobactéries pendant 48 h sur milieu gélosé. Puis, les explants sont rincés 2 fois 3 min dans des milieux de culture avec l'antibiotique de sélection, néomycine jusqu'à la



formation de cals. Les différents milieux de culture sont dans l'ordre : composition des milieux de préculture, de co-culture, de régénération, de callogénèse et d'enracinement (S. Ochatt, communication personnelle).

Les régénérants résistants sont repiqués sur le milieu de sélection et les plantes enracinées sont transférées en serre. Les plantes transformées sont analysées par PCR de façon à prouver l'intégration de l'ADN-T dans l'ADN génomique de la plante. Deux générations de plantes transformées sont générées avant les premières analyses moléculaires.

## 5. TILLING ET ECOTILLING

### **5.1 Criblage de la population de mutants EMS par la méthode TILLING**

Le principe du crible TILLING est présenté dans le chapitre 2 (section 4.1.1.1 et Figure 2.3). La technique détaillée est disponible dans la publication Dalmais *et al.* 2008.

L'ADN génomique des 4000 individus de la population EMS est extrait et les ADN sont poolés par groupe de huit individus. Le criblage s'effectue par une technique de PCR nichée suivie d'une digestion par l'endonucléase Cell. Les amorces utilisées pour l'amplification du gène ont été dessinées à partir de la séquence génomique disponible dans la banque de clones BAC de *M. truncatula* (Sequencing *Medicago truncatula* at the University of Oklahoma, [www.genome.ou.edu/medicago.html](http://www.genome.ou.edu/medicago.html)). La première PCR a été réalisée dans un mélange réactionnel de 11µl contenant 5µl d'ADN (comme décrit dans la section 2.5). La deuxième PCR est réalisée dans un mélange réactionnel de 11µl contenant 2µl de produit d'amplification de la première PCR des amorces internes marquées par des fluorochromes IRD-700 et IRD-800 (respectivement sens et antisens). Cette deuxième PCR est suivie d'une dénaturation de 10 minutes à 99°C, puis de 1 minute à 72°C et 70 cycles de 20s avec une diminution de 0,3°C par cycle. Cette étape permet la formation des hétéroduplexes entre les brins d'ADN sauvages et mutés. Les 11µl des produits de la deuxième amplification ont par la suite été digérés par 0,2µL de l'endonucléase Cell en présence de 2µl de tampon 10X (HEPES 0,1M pH 7,5 ; MgSO<sub>4</sub> 0,1M ; Triton X-100 0,02%). La réaction a été arrêtée par une solution d'EDTA. Afin d'éliminer les sels et les amorces, les échantillons ont été filtrés sur colonnes sephadex G50. 10µl des produits de filtration ont été séchés et repris dans 3 µl de tampon de charge. Avant le dépôt, les échantillons ont été dénaturés 2 minutes 30 secondes à 96°C. La migration a été réalisée sur gel d'acrylamide dénaturant (Acrylamide KBPlus 6.5%, 25cm de longueur et 0.25mm d'épaisseur) dans un séquenceur Li-Cor 4300 DNA Analyzer (migration à 1500V pendant 2 heures). La visualisation de la fluorescence à 700 et 800 nm se fait à l'aide de faisceaux laser intégrés au séquenceur. La présence et la position des mutations sont visualisées sur le gel grâce à la taille des produits de digestion. Le gène muté ainsi identifié est ensuite séquencé pour vérifier la position exacte et la nature de la mutation. Entre chaque génération de plante (autofécondation et rétrocroisement), il est nécessaire de vérifier la présence de la mutation et son état d'hétérozygotie. Ce suivi est réalisé soit par méthode tilling (pois) ou par dCAPS (*M. truncatula*).





### **5.2 EcoTILLING**

Le protocole d'EcoTILLING est le même que celui du TILLING à la seule différence que l'ADN génomique des individus de la collection de référence de pois est mélangé avec l'ADN génomique de l'écotype de référence, Caméor (ratio 1:1). Ensuite, les étapes sont les mêmes que précédemment décrites pour le TILLING.

## **6. ANALYSES BIOINFORMATIQUES**

La majorité des analyses bioinformatiques ont été réalisées en utilisant les outils disponibles dans la suite logicielle wEMBOSS (Sarachu et Colet 2005, <http://wembooss.sourceforge.net/>)(disponible sur le serveur de l'unité) et les outils disponibles sur le site LEGOO (<http://www.legoo.org/>).



# **RÉFÉRENCES BIBLIOGRAPHIQUES**



## Références Bibliographiques

- Aida M, Ishida T, Tasaka M (1999) Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the CUP-SHAPED COTYLEDON and SHOOT MERISTEMLESS genes. *Development* 126: 1563-1570.
- Alemanno, L., Devic, M., Niemenak, N., Sanier, C., Guilleminot, J., Rio, M., Verdeil, J.L., Montoro, P. (2008) Characterization of leafy cotyledon1-like during embryogenesis in *Theobroma cacao* L. *Planta* 227(4):853-66.
- Allen SE (1974) *Chemical analysis of ecological materials*. New York: John Wiley and Sons.
- Ané JM, Zhu H, Frugoli J (2008) Recent advances in *Medicago truncatula* genomics. *International Journal of Plant Genomics* Article ID 256597, doi:10.1155/2008/256597
- Baniel, A., Bertrand, D., Lelion, A., Gueguen, J. (1998). Variability in protein composition of pea seed studied by FPLC and multidimensional analysis. *Crop Science* 38, 1568–1575.
- Barker, D. G., S. Bianchi, F. Blondon, Y. Dattée, G. Duc, S. Essad, P. Flament, P. Gallusci, G. Génier, P. Guy, X. Muel, J. Tourneur, J. Dénarié, and T. Huguet. (1990). *Medicago truncatula*, a model plant system for studying the molecular genetics of the Rhizobium-legume symbiosis. *Plant Mol. Biol. Rep.* 8:40-49.
- Bassuner, R., Manteuffel, R., Muntz, K., Puchel, M., Schmidt, P. and Weber, E. (1983) Analysis of in vivo and in vitro globulin formation during cotyledon development of field beans (*Vicia faba* L. var. minor). *Biochemie und Physiologie der Pflanzen* 178(8): 665-684.
- Bäumlein, H., Boerjan, W., Nagy, I., Bassfüner, R., Van Montagu, M., Inzé, D. and Wobus, U. (1991) A novel seed protein gene from *Vicia faba* is developmentally regulated in transgenic tobacco and Arabidopsis plants. *Molecular and General Genetics* 225(3): 459-467.
- Bäumlein, H., Misera, S., Luerssen, H., Kolle, K., Horstmann, C., Wobus, U., Müller, A. (1994) The FUS3 gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. *Plant J.* 6: 367-387.
- Bäumlein, H., Nagy, I., Villarroel, R., Inzé, D., Wobus, U. (1992) Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. *Plant J.* 2(2):233-9.
- Bäumlein, H., Wobus, U., Pustell, J., Kafatos, F.C. (1986) The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucleic Acids Res.* 14(6):2707-20.
- Beach, L.R., Spencer, D., Randall, P.J., Higgins, T.J. (1985) Transcriptional and post-transcriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. *Nucleic Acids Res.* 13(3):999-1013.



- Benedito VA, Torres-Jerez I, Murray JD, Andrianakaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK. (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* .
- Bensmihen S, Giraudat J, Parcy F (2005) Characterization of three homologous basic leucine zipper transcription factors (bZIP) of the ABI5 family during *Arabidopsis thaliana* embryo maturation. *J Exp Bot* 56: 597-603
- Bensmihen, S., Rippa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F., Giraudat, J., Parcy, F. (2002) The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. *Plant Cell* 14(6):1391-403.
- Berger F. (2003). Endosperm: the crossroad of seed development. *Curr Opin Plant Biol.* 6(1):42-50.
- Bhattacharyya MK, Smith AM, Ellis THN, Hedley C and Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*, 60(1): 115-122.
- Bobb, A.J., Chern, M.S., Bustos, M.M. (1997) Conserved RY-repeats mediate transactivation of seed-specific promoters by the developmental regulator PvALF. *Nucleic Acids Res.* 25(3):641-7.
- Bobb, A.J., Eiben, H.G., Bustos, M.M. (1995) PvAlf, an embryo-specific acidic transcriptional activator enhances gene expression from phaseolin and phytohemagglutinin promoters. *Plant J.* 8(3):331-43.
- Borisjuk, L., Nguyen, T.H., Neuberger, T., Rutten, T., Tschiersch, H. et al (2005) Gradients of lipid storage, photosynthesis and plastid differentiation in developing soybean seeds. *New Phytol.* 167(3):761-76.
- Bouchez D. (2004). L'espèce modèle *Arabidopsis thaliana*. In : La génomique en biologie végétale. JF Briat & JF Morot-Gaudry, Eds. INRA Editions. pp 289-301
- Boulter D, Evans MI, Ellis RJ, Shirsat A, Gatehouse JA, Croy RRD (1987) Differential gene expression in the development of *Pisum sativum*. *Plant Physiol Biochem* 25: 283-89
- Boulter, D. (University of Durham (UK). Department of Botany); Evans, I.A.; Ellis, J.R.; Shirsat, A.; Gatehouse, J.A.; Croy, R.R.D. (1987) Differential gene expression in the development of *Pisum sativum* [transcriptional control, post-transcriptional control, legumin, vicilin] *Plant Physiology and Biochemistry.* 25(3) p. 283-289
- Bouton S, Viau L, Lelievre E, Limami A (2005) A gene encoding a protein with a proline-rich domain (MtPPRD1), revealed by suppressive subtractive hybridization (SSH), is specifically expressed in the *Medicago truncatula* embryo axis during germination. *J Exp Bot* 56: 825-832
- Bradford MM. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal Biochem* 72:248-54.
- Braybrook, S.A., Stone, S.L., Park, S., Bui, A.Q., Le, B.H., Fischer, R.L., Goldberg, R.B., Harada, J.J. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc Natl Acad Sci USA.* 103(9):3468-73.
- Britten, R.J. and Davidson, E.H. (1969) *Gene Regulation for Higher Cells: A Theory.* Science 1969: 349-357.
- Brocard-Gifford IM, Lynch TJ, Finkelstein RR (2003) Regulatory networks in seeds integrating developmental, abscisic acid, sugar, and light signaling. *Plant Physiol* 131:78-92





- Buitink J, Leger JJ, Guisle I, Vu BL, Wuillème S, Lamirault G, Le Bars A, Le Meur N, Becker A, Küster H, Leprince O. (2006). Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. *Plant J.* 47(5):735-50.
- Busk, P.K. and Pagès, M. (1998) Regulation of abscisic acid-induced transcription. *Plant Mol. Biol.* 37(3): 425-435.
- Caldo RA, Nettleton D, Wise RP. (2004) Interaction-dependent gene expression in Mla-specified response to barley powdery mildew. *Plant Cell.*16(9):2514-28.
- Casson, S.A. and Lindsey, K. (2006) The turnip mutant of *Arabidopsis* reveals that LEAFY COTYLEDON1 expression mediates the effects of auxin and sugars to promote embryonic cell identity. *Plant Physiol.* 142(2):526-41.
- Castle, L.A., Meinke, D.W. (1994) A FUSCA gene of *Arabidopsis* encodes a novel protein essential for plant development. *Plant Cell* 6(1):25-41.
- Cernac A, Benning C (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. *Plant J.* 40: 575-85
- Chandrasekharan MB, Bishop KJ, Hall TC (2003) Module-specific regulation of the beta-phaseolin promoter during embryogenesis. *Plant Journal* 33: 853-866
- Chang S, Pur Year J, Carney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep* 11:113-116
- Che P, Gingerich DJ, Lall S, Howell SH. (2002) Global and hormone-induced gene expression changes during shoot development in *Arabidopsis*. *Plant Cell.* 14(11):2771-85.
- Chen W, Provar NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan,S, Zou G, Whitman SA (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14: 559-574
- Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S. (2002) Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*. *Plant Physiol.* 129(2):661-77.
- Chern, M.S., Bobb, A.J., Bustos, M.M. (1996b) The regulator of MAT2 (ROM2) protein binds to early maturation promoters and represses PvALF-activated transcription. *Plant Cell.* 8(2):305-21.
- Chern, M.S., Eiben, H.G., Bustos, M.M. (1996a) The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. *Plant J.* 10(1):135-48.
- Cho Y, Fernandes J, Kim SH, Walbot V. (2002) Gene-expression profile comparisons distinguish seven organs of maize. *Genome Biol.* 3(9): research0045.
- Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR. (2004). A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics.* 166(3):1463-502.



- Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR. (2004). Estimating genome conservation between crop and model legume species. *Proc Natl Acad Sci USA*. 101(43):15289-94.
- Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S. (2001) High-throughput screening for induced point mutations. *Plant Physiol*. 126(2):480-4.
- Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J*. 37(5):778-86.
- Cook RD (1999) *Medicago truncatula* – A model in the making! *Current Opinion in Plant Biology* 2: 301-304
- Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S, Guimil S, Dunn M, Luginbühl P, Ellero C, Goff SA, Glazebrook J. (2003) A network of rice genes associated with stress response and seed development. *Proc Natl Acad Sci USA*. 100(8):4945-50.
- Curaba, J., Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M., Vachon, G. (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. *Plant Physiol*. 136(3):3660-9.
- Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root and shoot specific genes. *The Plant Journal* 38: 366-379.
- D**almais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savoie V, Aubert G, Brunaud V, de Oliveira Y, Guichard C, Thompson R, Bendahmane (2008) A. UTILdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol*. 9(2):R43.
- de Folter S, Busscher J, Colombo L, Losa A, Angenent G (2004) Transcript profiling of transcription factor genes during silique development in Arabidopsis. *Plant Mol Biol* 56: 351-366
- de Pater, S., Pham, K., Chua, N.H., Memelink, J., Kijne, J. (1993) A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. *Plant Cell*. 5(8):877-86.
- DeFrancesco L, Perkel JM (2001) In search of genomic variation: a wealth of technologies exists to find elusive genetic polymorphisms. *Scientist*, 15, 24.
- Dellaporta S. L., Wood, and J. B. Hicks. (1983). A plant DNA miniprep: version II. *Plant Molecular Biology Reporter* 1(14):19-21.
- d'Erfurth I, Cosson V, Eschstruth A, Lucas H, Kondorosi A, Ratet P. (2003) Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant J*. 34(1):95-106.
- Dhaubhadel S, McGarvey BD, Williams R, Gijzen M. (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. *Plant Mol Biol*. 53(6):733-43.
- Diaz I, Vicente-Carbajosa J, Abraham Z, Martínez M, Isabel-La Moneda I, Carbonero P (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant J* 29: 453-64



- Djemel N, Guedon D, Lechevalier A, Salon C, Miquel M, Prosperi JM, Rochat C, Boutin JP. (2005). Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol Biochem.* 43(6):557-66.
- Druka A, Muehlbauer G, Druka I, Caldo R, Baumann U, Rostoks N, Schreiber A, Wise R, Close T, Kleinhofs A, Graner A, Schulman A, Langridge P, Sato K, Hayes P, McNicol J, Marshall D, Waugh R (2006) An atlas of gene expression from seed to seed through barley development. *Funct Integr Genomics* 6:202-11
- E**dwards, D., Murray, J.A., Smith, A.G. (1998) Multiple genes encoding the conserved CCAAT-box transcription factor complex are expressed in *Arabidopsis*. *Plant Physiol.* 117(3):1015-22.
- Edwards K, Johnstone C, and Thompson C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res.* 19(6): 1349.
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 99: 10611-10616
- El Yahyaoui F, Küster H, Ben Amor B, Hohnjec N, Pühler A, Becker A, Gouzy J, Vernié T, Gough C, Niebel A, Godiard L, Gamas P. (2004). Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. *Plant Physiol.* 136(2):3159-76.
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY super family of plant transcription factors. *Trends Plant Sci* 5: 199-206
- Evans, I.M., Gatehouse, J.A., Croy, R.R.D. and Boulter, D. (1984) Regulation of the transcription of storage-protein mRNA in nuclei isolated from developing pea (*Pisum sativum* L.) cotyledons. *Planta* 160(6): 559-568.
- Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K. and Rask, L. (1999) Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol Biol.* 40(4):699-709.
- Ezcurra, I., Wycliffe, P., Nehlin, L., Ellerström, M., Rask, L. (2000) Transactivation of the *Brassica napus* napin promoter by ABI3 requires interaction of the conserved B2 and B3 domains of ABI3 with different cis-elements: B2 mediates activation through an ABRE, whereas B3 interacts with an RY/G-box. *Plant J.* 24(1):57-66.
- F**ait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiology* 142: 839-854
- Fambrini M, Durante C, Cionini G, Geri C, Giorgetti L, Michelotti V, Salvini M, Pugliesi C. (2006) Characterization of LEAFY COTYLEDON1-LIKE gene in *Helianthus annuus* and its relationship with zygotic and somatic embryogenesis. *Dev Genes Evol.* 216(5):253-64.
- Finkelstein RR, Lynch TJ. (2000) The *Arabidopsis* abscisic acid response gene ABI5 encodes a basic leucine zipper transcription factor. *Plant Cell.* 12(4):599-609.



- Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998) The Arabidopsis abscisic acid response locus ABI4 encodes an APETALA 2 domain protein. *Plant Cell* 10: 1043-1054
- Finkelstein R.R. and Gibson, S.I. (2002) ABA and sugar interactions regulating development: cross-talk or voices in a crowd. *Curr Opin Plant Biol.* 5(1):26-32.
- Flanagan CA, Ma H (1994) Spatially and temporally regulated expression of the MADS-box gene AGL2 in wild-type and mutant Arabidopsis flowers. *Plant Mol Biol.* 26: 581-95
- Focks N., Benning C. (1998) Wrinkled1: A novel low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate. *Plant Physiol.* 118: 91-101.
- G**aj, M.D., Zhang, S., Harada, J.J. and Lemaux, P.G. (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. *Planta* 222(6):977-88.
- Gallardo K, Firnhaber C, Zuber H, Hélicher D, Belghazi M, Henry C, Küster H, Thompson R. (2007). A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol Cell Proteomics.* (12):2165-79.
- Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D.(2001). Proteomic analysis of arabidopsis seed germination and priming. *Plant Physiol.* 126(2):835-48.
- Gallardo K, Job C, Groot SP, Puype M, Demol H, Vandekerckhove J, Job D. (2002). Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol.* 129(2):823-37.
- Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J. (2003). Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* 133(2):664-82.
- Garcia D, Fitz Gerald JN, Berger F (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *Plant Cell* 17: 52-60
- Gatehouse JA, Evans IM, Bown D, Croy RR, Boulter D. (1982) Control of storage-protein synthesis during seed development in pea (*Pisum sativum* L.) *Biochem J.* 208(1):119-27.
- Gatehouse JA, Evans IM, Bown D, Croy RR, Boulter D. (1986) Differential Expression of Genes During Legume Seed Development. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 314, No. 1166, Differential Gene Expression in the Regulation of Plant Growth and Development pp. 367-384
- Gatehouse JA, Evans IM, Croy RRD, Boulter D (1986) Differential expression of genes during legume seed development. *Philosophical Transactions of the Royal Society of London. Biological Sciences* 314: 367-384
- Gazzarrini, S., Tsuchiya, Y., Luma, S., Okamoto, M., McCourt, P. (2004) The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones Gibberellin and Abscisic Acid. *Dev. Cell* 7: 373-385.
- Gibson, S.I. (2005) Control of plant development and gene expression by sugar signaling. *Curr Opin Plant Biol.* 8(1):93-102.





- Giraudat, J., Hauge, B.M., Valon, C., Smalle, J., Parcy, F., Goodman, H.M. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4(10):1251-61.
- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J. (2000) Microarray analysis of developing Arabidopsis seeds. *Plant Physiol.* 124(4):1570-81.
- Goda H, Shimada Y, Asami T, Fujioka S, Yoshida S. (2002) Microarray analysis of brassinosteroid-regulated genes in Arabidopsis. *Plant Physiol.* 130(3):1319-34.
- Goffard N, Weiller G. (2006) Extending MapMan: application to legume genome arrays. *Bioinformatics.* 22(23):2958-9.
- Goffard N, Weiller G. GeneBins: (2007) a database for classifying gene expression data, with application to plant genome arrays. *BMC Bioinformatics.* 8:87.
- Graham MA, Silverstein KA, Cannon SB, VandenBosch KA (2004) Computational identification and characterization of novel genes from legumes. *Plant Physiol* 135:1179-97
- Graham PH, Vance CP. (2003) Legumes: importance and constraints to greater use. *Plant Physiol.* 131(3):872-7.
- Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S. (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics.* 164(2):731-40.
- Greenwood, J.S. and Chrispeels, M.J. (1985) Correct Targeting of the Bean Storage Protein Phaseolin in the Seeds of Transformed Tobacco. *Plant Physiol.* 79(1): 65–71.
- Gundry CN, Vandersteen JG, Reed GH et al. (2003) Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clinical Chemistry*, 49, 396–406.
- Gusmaroli G, Tonelli C, Mantovani R. (2001) Regulation of the CCAAT-Binding NF-Y subunits in Arabidopsis thaliana. *Gene.* 264(2):173-85.
- Gutierrez, L., Conejero, G., Castelain, M., Guénin, S., Verdeil, J.L., Thomasset, B., Van Wuytswinkel, O. (2006) Identification of new gene expression regulators specifically expressed during plant seed maturation. *J Exp Bot.* 57(9):1919-32.
- Gutierrez, L., Van Wuytswinkel, O., Castelain, M., Bellini, C. (2007) Combined networks regulating seed maturation. *Trends Plant Sci.* 12(7):294-300.
- H**annon GJ. (2002) RNA interference. *Nature.* 418(6894):244-51.
- Harada JJ (2001) Role of Arabidopsis LEAFY COTYLEDON genes in seed development. *Journal of Plant Physiology* 158(4), p 405-409.
- Harada, J.J., Barker, S.J. and Goldberg, R.B. (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell.* 1(4):415-25.
- Hardtke CS, Berleth T (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J* 17: 1405-11



- Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. (2000) Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. *Science*. 290(5499):2110-3.
- Helliwell C, Waterhouse P. (2003) Constructs and methods for high-throughput gene silencing in plants. *Methods*. 30(4):289-95.
- Henderson, J.T., Li, H.C., Rider, S.D., Mordhorst, A.P., Romero-Severson, J., Cheng, J.C., Robey, J., Sung, Z.R., De Vries, S.C., Ogas, J. (2004) PICKLE acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. *Plant Physiol*. 134(3):995-1005.
- Hernández-Sebastià, C., Marsolais, F., Saravitz, C., Israel, D., Dewey, R.E., Huber, S.C. (2005) Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low- and high-protein soybean lines. *J Exp Bot*. 56(417):1951-63.
- Hiratsu K, Matsui K, Koyama T, Ohme-Takagi M. (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. *Plant J*. 34(5):733-9.
- Hobo, T., Kowyama, Y., Hattori, T. (1999) A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. *Proc Natl Acad Sci USA*. 96(26):15348-53.
- Hoefgen, R; Nikiforova, VJ (2008) Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiologia Plantarum*, 132 (2): 190-198
- Holmes P, Goffard N, Weiller GF, Rolfe BG, Imin N. (2008) Transcriptional profiling of *Medicago truncatula* meristematic root cells. *BMC Plant Biol*. 8:21.
- Huijser C, Kortstee A, Pego J, Weisbeek P, Wisman E, Smeekens S. (2000) The Arabidopsis SUCROSE UNCOUPLED-6 gene is identical to ABSCISIC ACID INSENSITIVE-4: involvement of abscisic acid in sugar responses. *Plant J*. 23(5):577-85.
- Hunter BG, Beatty MK, Singletary GW, Hamaker BR, Dilkes BP, Larkins BA, Jung R. (2002) Maize opaque endosperm mutations create extensive changes in patterns of gene expression. *Plant Cell*. 14(10):2591-612.
- Huntley, R.P. and Murray, J.A.H. (1999) The plant cell cycle. *Current Opinion in Plant Biology* 2(6): 440-446.
- J**ürgens G. (2001). Apical-basal pattern formation in Arabidopsis embryogenesis. *EMBO J*. 20(14):3609-16. Review.
- K**agaya Y, Okuda R, Ban A, Toyoshima R, Tsutsumida K, Usui H, Yamamoto A, Hattori T. (2005b) Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in Arabidopsis. *Plant Cell Physiol*. 46(2):300-11.
- Kagaya Y, Toyoshima R, Okuda R, Usui H, Yamamoto A, Hattori T (2005a) LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol* 46: 399-406
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 32(Database issue):D277-80.



- Kawagoe, Y. and Mura, N. (1996) A novel basic region/helix-loop-helix protein binds to a G-box motif CACGTG of the bean seed storage protein  $\beta$ -phaseolin gene. *Plant Science* 116(1): 47-57.
- Kawagoe, Y., Campbell, B.R., Murai, N. (1994) Synergism between CACGTG (G-box) and CACCTG cis-elements is required for activation of the bean seed storage protein beta-phaseolin gene. *Plant J.* 5(6):885-90.
- Keith, K., Kraml, M., Dengler, N.G., McCourt, P.(1994) *fusca3*: A Heterochronic Mutation Affecting Late Embryo Development in Arabidopsis. *Plant Cell* 6(5):589-600.
- Kim IS, Sinha S, de Crombrughe B, Maity SN. (1996) Determination of functional domains in the C subunit of the CCAAT-binding factor (CBF) necessary for formation of a CBF-DNA complex: CBF-B interacts simultaneously with both the CBF-A and CBF-C subunits to form a heterotrimeric CBF molecule. *Mol Cell Biol.* 16(8):4003-13.
- Koch K. (2004) Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr Opin Plant Biol.* 7(3):235-46.
- Kroj T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in Arabidopsis. *Development* 130: 6065-73
- Künne C, Lange M, Funke T, Miede H, Thiel T, Grosse I, Scholz U. (2005) CR-EST: a resource for crop ESTs. *Nucleic Acids Res.* 33(Database issue):D619-21.
- Kushiro, T., Nambara, E. and McCourt P. (2003) Hormone evolution: The key to signalling. *Nature* 422(6928):122.
- Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15: 5-18
- Ladin, B.F., Tierney, M.L., Meinke, D.W., Hosángadi, P., Veith, M., and Beachy, R.N. (1987) Developmental Regulation of  $\beta$ -Conglycinin in Soybean Axes and Cotyledons. *Plant Physiol.* 84(1): 35-41.
- Lara P, Oñate-Sánchez L, Abraham Z, Ferrándiz C, Díaz I, Carbonero P, Vicente-Carbajosa J (2003) Synergistic activation of seed storage protein gene expression in Arabidopsis by ABI3 and two bZIPs related to OPAQUE2. *J Biol Chem* 278: 21003-11
- Larkin, P.J. and Scowcroft, W.R. (1981) Somaclonal variation: a novel source of variability from cell culture for plant improvement. *Theor. Appl. Genet.* 60, 197-214.
- Lee H, Fischer RL, Goldberg RB, Harada JJ. (2003) Arabidopsis LEAFY COTYLEDON1 represents a functionally specialized subunit of the CCAAT binding transcription factor. *Proc Natl Acad Sci USA.* 100(4):2152-6.
- Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. (2006) Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol.* 57:405-30.
- Li G, Chandrasekharan MB, Wolffe AP, Hall TC (2001) Chromatin structure and phaseolin gene regulation. *Plant Mol Biol* 46: 121-9
- Li X, Lassner M, Zhang Y. (2002) Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. *Comp Funct Genomics.* 3(2):158-60.



- Long JA, moan EI, Medford JI, Barton MK (1996) A member of the KNOTTED class of homeodomain proteins encoded by the STM gene of Arabidopsis. *Nature* 404: 66-69
- Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93: 1195-1205
- Lu P, Porat R, Nadeau JA, O'Neill SD (1996) Identification of a meristem L1 layer-specific gene in Arabidopsis that is expressed during embryonic pattern formation and defines a new class of homeobox genes. *Plant Cell* 8: 2155-68
- M**aity, S.N. and De Crombrughe, B. (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci.* 23(5):174-8.
- Mansfield, S.G., and Briarty, L.G. (1990a) Development of the free-nuclear endosperm in Arabidopsis thaliana. *Arabidopsis Inf. Serv.* 27:53-64.
- Mansfield, S.G., and Briarty, L.G. (1990b) Endosperm cellularization in Arabidopsis thaliana. *Arabidopsis Inf. Serv.* 27:65-72.
- Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG. (2001). Establishment of a root proteome reference map for the model legume Medicago truncatula using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics.* 1(11):1424-40.
- Mayer KF, Schoof H, Haecker A, Lenhard M, Jurgens G, Laux T (1998) Role of WUSCHEL in regulating stem cell fate in Arabidopsis shoot meristem. *Cell* 95: 805-815
- Mayer, U., Torres Ruiz, R. A., Berleth, T., Misera, S. and Jürgens, G. (1991). Mutations affecting body organization in the Arabidopsis embryo. *Nature* 353,402 -407
- McCallum CM, Comai L, Greene EA, Henikoff S. (2000) Targeted screening for induced mutations. *Nat Biotechnol.* 18(4):455-7.
- McCallum CM, Comai L, Greene EA, Henikoff S. (2000) Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123(2):439-42.
- 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(5):895-905.
- McCourt, P. (1999) Analysis of hormone signalling. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 219-243.
- Meinke DW, Chen J, Beachy RN (1981) Expression of storage-protein genes during soybean seed development. *Planta* 153: 130-139
- Meinke DW, Franzmann LH, Nickle TC, Yeung EC. (1994) Leafy Cotyledon Mutants of Arabidopsis. *Plant Cell.* 6(8):1049-1064.
- Meinke, D.W. (1992) A Homoeotic Mutant of Arabidopsis thaliana with Leafy Cotyledons. *Science* 258(5088):1647-1650.





- Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P. (1998) An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamins-box of a native B-hordein promoter in barley endosperm. *Plant J.* 16(1):53-62.
- Menges M, Hennig L, Gruissem W, Murray JA. (2002) Cell cycle-regulated gene expression in Arabidopsis. *J Biol Chem.* 277(44):41987-2002.
- Mönke, G., Altschmied, L., Tewes, A., Reidt, W., Mock, H.P., Bäumlein, H., Conrad, U. (2004) Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. *Planta* 219(1):158-66.
- Morandi D, Prado E, Sagan M, Duc G. (2005) Characterisation of new symbiotic *Medicago truncatula* (Gaertn.) mutants, and phenotypic or genotypic complementary information on previously described mutants. *Mycorrhiza.* 15(4):283-9.
- Moseyko N, Zhu T, Chang HS, Wang X, Feldman LJ. (2002) Transcription profiling of the early gravitropic response in Arabidopsis using high-density oligonucleotide probe microarrays. *Plant Physiol.* 130(2):720-8.
- Mounet, F; Lemaire-Chamley, M; Maucourt, M; Cabasson, C; Giraudel, JL et al. (2007) Quantitative metabolic profiles of tomato flesh and seeds during fruit development: complementary analysis with ANN and PCA. *Metabolomics*, 3(3): 273-288.
- Munier-Jolain N.G., Carrouée B. (2003). Quelle place pour le pois dans une agriculture respectueuse de l'environnement : Argumentaire agri-environnemental. *Cahiers Agriculture*, 12, 111-120.
- Munier-Jolain NG, Ney B. (1998) Seed growth rate in grain legumes. II: Seed growth rate depends on cotyledon cell number, *J. Exp. Bot.* 49 1971-1976.
- Nakamura, S., Lynch, T.J., Finkelstein, R.R. (2001) Physical interactions between ABA response loci of Arabidopsis. *Plant J.* 26(6): 627-35.
- Nambara, E. and Marion-Poll, A. (2003) ABA action and interactions in seeds. *Trends Plant Sci.* 8(5):213-7.
- Needleman SB, Wunsch CD. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* 48(3):443-53.
- Ng PC, Henikoff S. (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31(13):3812-4.
- Ng, D.W. and Hall, T.C. (2008) PvALF and FUS3 activate expression from the phaseolin promoter by different mechanisms. *Plant Mol Biol.* 66(3):233-44.
- Ng, D.W., Chandrasekharan, M.B., Hall, T.C. (2006) Ordered histone modifications are associated with transcriptional poising and activation of the phaseolin promoter. *Plant Cell* 18(1):119-32.
- Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallan, B.J., Fischer, R.L., Sims, T.L., Drews, G.N. and Goldberg, R.B. (1989) Characterization of the glycinin gene family in soybean. *Plant Cell.* 1(3):313-28.



- O**'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007-21.
- Ogas, J., Kaufmann, S., Henderson, J., Somerville, C. (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. *Proc Natl Acad Sci USA*. 96(24):13839-44.
- Ohto, M.A., Fischer, R.L., Goldberg, R.B., Nakamura, K., Harada, J.J. (2005) Control of seed mass by APETALA2. *Proc Natl Acad Sci USA*. 102(8):3123-8.
- P**ang, P.P., Pruitt, R.E. and Meyerowitz, E.M. (1988) Molecular cloning, genomic organization, expression and evolution of 12S seed storage protein genes of Arabidopsis thaliana. *Plant Mol. Biol.* 11(6): 805-820.
- Parcy F, Valon C, Kohara A, Miséra S, Giraudat J (1997) The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of Arabidopsis seed development. *Plant Cell* 9: 1265-77
- Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J (1995) Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell* 6: 1567-82
- Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., Giraudat, J. (1994) Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell*. 6(11):1567-82.
- Penfield S, Li Y, Gilday AD, Graham S, Graham IA. (2006) Arabidopsis ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *Plant Cell*. 18(8):1887-99.
- Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M. (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume Lotus japonicus. *Plant Physiol*. 131(3):866-71.
- Pla, M., Vilardell, J., Guiltinan, M.J., Marcotte, W.R., Niogret, M.F., Quatrano, R.S., Pagès M. (1993) The cis-regulatory element CCACGTGG is involved in ABA and water-stress responses of the maize gene rab28. *Plant Mol Biol*. 21(2):259-66.
- Potokina E, Prasad M, Malysheva L, Röder MS, Graner A. (2006) Expression genetics and haplotype analysis reveal cis regulation of serine carboxypeptidase I (Cxp1), a candidate gene for malting quality in barley (*Hordeum vulgare* L.). *Funct Integr Genomics*. 6(1):25-35.
- PROLEA : <http://www.prolea.com/>
- Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* 80: 847-57
- Q**u, L.J. and Zhu, Y.X. (2006) Transcription factor families in Arabidopsis: major progress and outstanding issues for future research. *Current Opinion in Plant Biology* 9(5): 544-549.



- R** Foundation for Statistical Computing. <http://www.R-project.org/> Editor-in-Chief: John Fox, Department of Sociology, McMaster University, Ontario, Canada
- Rajjou L, Gallardo K, Debeaujon I, Vandekerekhove J, Job C, Job D (2004) The Effect of  $\alpha$ -amanitin on the Arabidopsis Seed Proteome Highlights the Distinct Roles of stored and Neosynthesized mRNAs during germination. *Plant Physiology* 134: 1598-1613
- Ramakers C, Ruijter JM, Lekanne Deprez RH, Moormana AFM (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* 339: 62-66
- Reidt W, Wohlfarth T, Ellerström M, Czihal A, Tewes A, Ezcurra I, Rask L, Bäumlein H (2000) Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J* 21: 401-8.
- Reidt, W., Ellerström, M., Kölle, K., Tewes, A., Tiedemann, J., Altschmied, L. and Bäumlein, H. (2001) FUS3-dependent gene regulation during late embryogenesis. *Journal of Plant Physiology* 158(4): 411-418.
- Repetto O, Rogniaux H, Firmhaber C, Zuber H, Küster H, Larré C, Thompson R, Gallardo K. (2008) Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. *Plant J.* (in press).
- Rider, S.D., 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. *Plant J.* 35(1):33-43.
- Riechmann JL (2002) Transcriptional regulation: a genomic overview (September, 30, 2002). In the Arabidopsis Book (Sommerville CR and Meyerowitz EM eds), Rockville, MD: American Society of Plant Biologists (<http://www.aspb.org/publications/Arabidopsis/>)
- Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. *Current Opinion in Plant Biology* 3: 423-434
- Rolland, F., Baena-Gonzalez, E. and Sheen, J. (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol.* 57:675-709.
- Rolletschek, H., Hosein, F., Miranda, M., Heim, U., Götz, K.P., Schlereth, A., Borisjuk, L., Saalbach, I., Wobus, U. and Weber, H. (2005b) Ectopic expression of an amino acid transporter (VfAAP1) in seeds of *Vicia narbonensis* and pea increases storage proteins. *Plant Physiol.* 137(4):1236-49.
- Rolletschek, H., Koch, K., Wobus, U. and Borisjuk, L. (2005a) Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. *Plant J.* 42(1):69-83.
- Rook F, Gerrits N, Kortstee A, van Kampen M, Borrias M, Weisbeek P, Smeeckens S (1998) Sucrose-specific signalling represses translation of the Arabidopsis ATB2 bZIP transcription factor gene. *Plant J* 15: 253-63
- Rossel JB, Wilson IW, Pogson BJ. (2002) Global changes in gene expression in response to high light in Arabidopsis. *Plant Physiol.* 130(3):1109-20.
- Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* 14: 1191-206



- Saiki, R. K., Scharf, S., Faloona, F. Mullis, K. B., Horn, G. T., Erlich, H. A. Arnheim, N. (1985). Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354.
- Salanoubat M, Lemcke K, Rieger M, Ansorge W, Unseld M, Fartmann B, Valle G, Blocker H, Perez-Alonso M, Obermaier B, et al (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408: 820-822
- Samac DA, Tesfaye M, Dornbusch M, Saruul P, Temple SJ. (2004) A comparison of constitutive promoters for expression of transgenes in alfalfa (*Medicago sativa*). *Transgenic Res.* 13(4):349-61.
- Santos Mendoza M, Dubreucq B, Miquel M, Caboche M, Lepiniec L (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Letter* 579: 4666-70
- Santos-Mendoza M, Dubreucq B, Baud S, Parcy F., Caboche M., Lepiniec L. (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. *Plant J.* 54: 608-620
- Sarachu M, Colet M. (2005). wEMBOSS: a web interface for EMBOSS. *Bioinformatics.* 21(4):540-1.
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet.* 37(5):501-6.
- Schmid M, Uhlenhaut NH, Godard F, Demar M, Bressan R, Weigel D, Lohmann JU. (2003) Dissection of floral induction pathways using global expression analysis. *Development.* 130(24):6001-12.
- Scholte M, d’Erfurth I, Rippa S, Mondy S, Cosson V, Durand P, Breda C, Trinh H, Rodriguez-Llorente I, Kondorosi E et al. (2002) T-DNA tagging in the model legume *Medicago truncatula* allows efficient gene discovery. *Mol Breed* 10: 203–215
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ (2006) The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133: 251-61
- Shotwell, M.A., and Larkins, B.A. (1989) The molecular biology and biochemistry of seed storage proteins. In *The Biochemistry of Plants*, Vol. 15, A. Marcus, ed. (San Diego, CA: Academic Press), pp. 297–345.
- Shujun Chang<sup>1</sup>, Jeff Puryear<sup>1</sup> and John Cairney (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* Volume 11(2): 113-116
- Singh, K.B., Foley, R.C. and Oñate-Sánchez, L. (2002) Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology* 5(5): 430-436.
- Sinha S, Kim IS, Sohn KY, de Crombrughe B, Maity SN. (1996) Three classes of mutations in the A subunit of the CCAAT-binding factor CBF delineate functional domains involved in the three-step assembly of the CBF-DNA complex. *Mol Cell Biol.* 16(1):328-37.
- Slabas AR, Fordham-Skelton AP, Fletcher D, Martinez-Rivas JM, Swinhoe R, Croy RR, Evans IM (1994) Characterisation of cDNA and genomic clones encoding homologues of the 65 kDa regulatory subunit of protein phosphatase 2A in *Arabidopsis thaliana*. *Plant Mol Biol.* 26:1125-38
- Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat Biotechnol.* 23(1):75-81.





- Sozzani R, Maggio C, Varotto S, Canova S, Bergounioux C, Albani D, Cella R (2006) Interplay between Arabidopsis activating factors E2Fb and E2Fa in cell cycle progression and development. *Plant Physiol.* 140(4):1355-1366
- Stålberg, K., Ellerstöm, M., Ezcurra, I., Ablov, S., Rask, L. (1996) Disruption of an overlapping E-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. *Planta* 199(4):515-9.
- Statistica StatSoft, Inc. (2007). Electronic Statistics Textbook. Tulsa, StatSoft. WEB: <http://www.statsoft.com/textbook/stathome.html>.
- Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ. (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci USA.* 98(20):11806-11.
- Stone, S.L., Braybrook, S.A., Paula, S.L., Kwong, L.W., Meuser, J., Pelletier, J., Hsieh, T.F., Fischer, R.L., Goldberg, R.B., Harada, J.J. (2008) Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc Natl Acad Sci USA.* 105(8):3151-6.
- Sturn A, Quackenbush J, Trajanoski Z (2002) Genesis: cluster analysis of microarray data. *Bioinformatics* 18: 207-208
- Sun, T. and Gubler, F. (2004) Molecular mechanism of gibberellin signalling in plants. *Annual Review of Plant Biology* 55: 197-223.
- Suzuki A, Wu CY, Washida H, Takaiwa F (1998) Rice MYB protein OSMYB5 specifically binds to the AACA motif conserved among promoters of genes for storage protein glutelin. *Plant Cell Physiol* 39: 555-559
- Suzuki M, Wang HH, McCarty DR (2007) Repression of the LEAFY COTYLEDON1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol* 143: 902-11
- Suzuki, M., Wang, H.H., McCarty, D.R. (2007) Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol.* 143(2):902-11.
- T**akatsuji H (1999) Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. *Plant Mol Biol* 39: 1073–1078
- Thompson JD, Higgins DG, Gibson TJ. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673-80.
- Thoquet P, Ghérardi M, Journet EP, Kereszt A, Ané JM, Prosperi JM, Huguet T. (2002). The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. *BMC Plant Biol.* 2:1
- Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S. (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4:12.



- To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., Parcy, F. (2006) A network of local and redundant gene regulation governs Arabidopsis seed maturation. *Plant Cell* 18(7):1642-51.
- Tsuchiya, Y., Nambara, E., Naito, S., McCourt, P. (2004) The FUS3 transcription factor functions through the epidermal regulator TTG1 during embryogenesis in Arabidopsis. *Plant J.* 37(1):73-81.
- Tsukagoshi H, Morikami A, Nakamura K (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc Natl Acad Sci USA* 104: 2543-7
- Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A., Nakamura, K. (2005) Analysis of a sugar response mutant of Arabidopsis identified a novel B3 domain protein that functions as an active transcriptional repressor. *Plant Physiol.* 138(2):675-85.
- Tzitzikas EN, Vincken JP, de Groot J, Gruppen H, Visser RG. (2006) Genetic variation in pea seed globulin composition. *J Agric Food Chem.* 54(2):425-33.
- U**dvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, Zhang JY, Benedito V, Hofer JM, Chueng F, Town CD (2007) Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol* 144: 538-49
- Unger E, Parsons RL, Schmidt RJ, Bowen B, Roth BA (2002) Dominant negative mutants of Opaque2 suppress transactivation of a 22-kD zein promoter by Opaque2 in maize endosperm cells. *Plant Cell* 5: 831-841
- Unger E, Parsons RL, Schmidt RJ, Bowen B, Roth BA. (1993) Dominant Negative Mutants of Opaque2 Suppress Transactivation of a 22-kD Zein Promoter by Opaque2 in Maize Endosperm Cells. *Plant Cell.* 1993 Aug;5(8):831-841.
- Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D, Scheible WR, Gibon Y, Morcuende R, Weicht D, Meyer S, Stitt M (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol* 138: 1195-204
- V**andesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* 3
- Verdier, J., Kakar, K., Gallardo, K., Le Signor, C., Aubert, G., Schlereth, A., Town, C.D., Udvardi, M.K., Thompson, R.D. (2008) Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol Biol.* 67(6):567-580.
- Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ. (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci USA.* 94(14):7685-90.
- Vicente-Carbajosa, J., Carbonero P. (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. *Int. J. Dev. Biol.* 49: 645-651.



- Vigeolas, H., Van Dongen, J.T., Waldeck, P., Huhn, D. and Geigenberger, P. (2003) Lipid storage metabolism is limited by the prevailing low oxygen concentrations within developing seeds of oilseed rape. *Plant Physiol.* 133(4):2048-60.
- Vigeolas, H; Chinoy, C; Zuther, E; Blessington, B; Geigenberger, P; Domoney, C (2008) Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds. *Plant Physiol.* 146(1):74-82.
- W**alling L, Drews GN, Goldberg RB (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. *Proc Natl Acad Sci USA* 83: 2123-2127
- Wang TL, Domoney C, Hedley CL, Casey R, Grusak MA (2003) Can we improve the nutritional quality of legume seeds? *Plant Physiol.* 2003 Mar;131(3):886-91.
- Watson BS, Asirvatham VS, Wang L, Sumner LW. (2003). Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol.* 131(3):1104-23.
- Weber H, Borisjuk L, Heim U, Buchner P, Wobus U (1995) Seed coat-associated invertases of fava bean control both unloading and storage functions: cloning of cDNAs and cell type-specific expression. *Plant Cell* 7: 1835-46.
- Weber H, Borisjuk L, Heim U, Sauer N, Wobus U. (1997) A role for sugar transporters during seed development: molecular characterization of a hexose and a sucrose carrier in fava bean seeds. *Plant Cell* 9(6):895-908.
- Weber H, Borisjuk L, Wobus U. (2005). Molecular physiology of legume seed development. *Annu Rev Plant Biol.* 56:253-79.
- Wellmer F, Riechmann JL, Alves-Ferreira M, Meyerowitz EM. (2004) Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *Plant Cell* 16(5):1314-26.
- Wesley SV, Helliwell CA, Smith NA, Wang MB, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbott D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* 27(6):581-90.
- West M, Yee KM, Danao J, Zimmerman JL, Fischer RL, Goldberg RB, Harada JJ. (1994) LEAFY COTYLEDON1 is an Essential Regulator of Late Embryogenesis and Cotyledon Identity in *Arabidopsis*. *Plant Cell.* 6(12):1731-1745.
- Wilm M. (2000). Mass spectrometric analysis of proteins. *Adv. Protein. Chem.* 54:1-30.
- Y**amamoto MP, Onodera Y, Touno SM, Takaiwa F (2006) Synergism between RPB1 Dof and RISBZ1 bZIP Activators in the Regulation of Rice Seed Expression Genes. *Plant Physiol* 141: 1694-1707.
- Yazawa K, Takahata K, Kamada H. (2004) Isolation of the gene encoding Carrot leafy cotyledon1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol Biochem.* 42(3):215-23.



**Z**hu T, Budworth P, Chen W, Provart N, Chang HS, Guimil S, Su W, Estes B, Zou G, Wang X. (2003) Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnol J.* 1(1):59-70.





# **ANNEXES**



### Real-time RT-PCR profiling of all annotated *Medicago truncatula* transcription factors during seed filling.

J. Verdier<sup>1</sup>, K. Gallardo<sup>1</sup>, C. Lesignor<sup>1</sup>, S. Djennane<sup>1</sup>, G. Aubert<sup>1</sup>, D. Hericher<sup>1</sup>, C. Firnhaber<sup>2</sup>, H. Kuester<sup>2</sup>, M. Crespi<sup>3</sup>, C. Town<sup>5</sup>, K. Kakar<sup>4</sup>, M. Udvardi<sup>4</sup>, R. Thompson<sup>1</sup>

<sup>1</sup>URLEG Dijon (France), <sup>2</sup>Bielefeld University (Germany), <sup>3</sup>CNRS Gif/Yvette (France), <sup>4</sup>MPI-MP Golm (Germany), <sup>5</sup>TIGR (USA).

Seed filling is a key process for seed quality. We are studying this process in *Medicago truncatula* at the proteomic and transcriptomic levels. Our initial results have defined a number of stages of development of the seed with the corresponding metabolic activities. These studies have recently been widened by corresponding analyses being carried out at the transcriptional level with the aid of microarray technology.

Bearing in mind the current limits in detection sensitivity imposed by microarray technology, we have elected to profile expression of a panel of *Medicago truncatula* transcription factors during seed development using real-time reverse transcription-PCR. The Max Planck Institute of Molecular Plant Physiology (MPI-MP) at Golm (Germany) has developed a resource for quantitative measurement of transcripts encoding all annotated transcription factors (TFs). This collection is subjected to transcript profiling during seed filling.

The stages of seed development chosen for this analysis are the same ones taken for the comparative transcriptomics and proteomics study described above (Gallardo *et al.* 2003 and unpublished results) and include the so-far identified major events in seed filling in *Medicago truncatula*.

The aim of this study is to identify genes encoding seed-specific transcription factors and to reveal their function in specific processes during seed development such as storage reserve accumulation. Once genes of interest have been identified, mutant alleles will be screened within a TILLING resource for *Medicago truncatula* established in our laboratory.



### Gene expression dynamics during *M. truncatula* seed formation

**J Verdier**<sup>1</sup>, K. Gallardo<sup>1</sup>, C. Firnhaber<sup>2</sup>, K. Kakar<sup>3</sup>, C Lesignor<sup>1</sup>, G. Aubert<sup>1</sup>, M. Udvardi<sup>3</sup>, H. Küster<sup>2</sup>, R. Thompson<sup>1</sup>

<sup>1</sup> INRA-URLEG Dijon (France)

<sup>2</sup> Bielefeld University (Germany)

<sup>3</sup> MPI-MP Golm (Germany)

Gene expression dynamics was investigated in developing seeds at the transcript level using microarrays representing 16,086 annotated genes and at the protein level by high-resolution 2D-gels. These studies have allowed us to generate profiles for proteins and transcripts during entire seed development, and in the component seed tissues: the integuments, endosperm and embryo. The supporting roles of integument and endosperm for storage compound accumulation within the embryo are clearly demonstrated by these profiles, as are the metabolic switches occurring during seed filling and in late maturation that prepare the seed for dormancy and subsequent germination. By comparing protein accumulation with that of transcripts, a number of genes can be classified according to their predominantly transcriptional or post-transcriptional regulation. Bearing in mind the current limits in detection sensitivity imposed by these technologies to profile lowly expressed transcription factors, 1400 *M. truncatula* genes encoding known and putative transcription factors were analysed by quantitative reverse transcription-PCR throughout seed development. Genes expressed during specific stages were identified, among them several potential regulators of storage functions. Mutant alleles for these genes will be screened within a TILLING resource for *M. truncatula* established in our laboratory.

#### Acknowledgements.

This research was supported by the FP6 EU project “Grain Legumes” (FOOD-CT-2004-506223) and the Regional Council of Burgundy.



### **Etude des facteurs de transcription lors du développement de la graine de *M. truncatula* : cinétiques d'expression et compartimentation tissulaire**

Jérôme Verdier<sup>1</sup>, Klementina Kakar<sup>2</sup>, Karine Gallardo<sup>1</sup>, Christine Le Signor<sup>1</sup>, Grégoire Aubert<sup>1</sup>, Chris Town<sup>3</sup>, Michael Udvardi<sup>2</sup>, Richard Thompson<sup>1</sup>

<sup>1</sup>INRA Unité Mixte de Recherches en Génétique et Ecophysiologie des légumineuses à graines (UMR-LEG) - Dijon (France) ; <sup>2</sup>Max-Planck Institut für Molekulare Pflanzenphysiologie - Potsdam (Germany) ; <sup>3</sup>The Institute for Genomic Research (TIGR) - Rockville, MD (USA)

La qualité nutritionnelle des graines des légumineuses constitue un enjeu économique important face aux besoins croissants en protéines végétales destinées à l'alimentation animale et humaine. Cependant, certains caractères, tels que la teneur en protéines des graines, peuvent être améliorés. Nous avons récemment montré qu'une majorité des protéines intervenant dans le développement de la graine étaient principalement soumises à une régulation de type transcriptionnelle, y compris les différentes protéines de réserve. Suite à cette constatation, nous avons réalisé les profils d'expression de plus de 700 facteurs de transcription (TFs) chez la légumineuse modèle *Medicago truncatula* en utilisant les outils de la plateforme de PCR quantitative haut débit. Sept stades clefs du développement de la graine ont été choisis allant de la fin de l'embryogénèse jusqu'à la dessiccation. 512 TFs ont été détectés au cours du développement de la graine dont 170 montrant une expression différentielle durant cette période. Sur la base de leur profil d'expression, ces derniers ont été répartis dans des groupes fonctionnels associés à des processus spécifiques du développement des graines : embryogénèse tardive, transition embryogénèse-remplissage, synthèse des vicilines, des légumine A, des légumine K et enfin dessiccation. Le profil d'expression dans les différents tissus de la graine (tégument, albumen et embryon) a également été entrepris pour les TFs associés aux processus de remplissage. Une comparaison entre notre analyse et des données transcriptomiques (16k gènes microarray, collaboration H. Kuester, Univ. Bielefeld) au cours du développement de la graine de *M. truncatula* (Gallardo *et al.*, in prep), nous a permis de proposer des hypothèses sur le rôle de certains TFs grâce à l'identification de leurs cibles potentielles. Des facteurs de transcription potentiellement importants pour le développement de la graine ont ainsi pu être identifiés. Leur fonction est en cours d'analyse au travers de l'étude de mutants obtenus en utilisant la ressource TILLING disponible au laboratoire.





### Real-time RT-PCR profiling of over 1400 *Medicago truncatula* transcription factors during seed filling.

Verdier J.<sup>1\*</sup>, Kakar K.<sup>2</sup>, Gallardo K.<sup>1</sup>, Lesignor C.<sup>1\*</sup>, Aubert G.<sup>1</sup>, Kuester H.<sup>3</sup>, Crespi M.<sup>4</sup>, Town C.<sup>5</sup>, Udvardi M.<sup>2</sup>, Thompson R.<sup>1\*</sup>

<sup>1</sup>INRA Unité de Recherches en Génétique et Ecophysiologie des légumineuses à graines (URLEG) - 21065 Dijon (France)

<sup>2</sup>Max-Planck-Institut für Molekulare Pflanzenphysiologie - 14424 Potsdam (Germany)

<sup>3</sup>Universität Bielefeld-Fakultät für Biologie - 33594 Bielefeld (Germany)

<sup>4</sup>CNRS Institut Sciences du végétal – 91198 Gif/Yvette (France)

<sup>5</sup>The Institute for Genomic Research (TIGR) - Rockville, MD 20850 (USA)

Seed filling is a key process for legume seed quality, the storage materials that constitute the value of seed crops being deposited during this critical period of seed development. We are using *Medicago truncatula*, developed in recent years as a model for legumes, to study seed development at the proteomic and transcriptomic levels, with the aid of microarrays. A number of stages of seed development with the corresponding metabolic networks have been defined. Bearing in mind the current limits in detection sensitivity imposed by these technologies for detecting regulatory factors, we have elected to profile expression of all annotated *M. truncatula* transcription factors during seed development by quantitative reverse transcription-PCR. Seven stages of seed development were chosen for this study. These correspond to embryogenesis, onset of seed filling, stages related to the massive synthesis of storage compounds, maturation and desiccation.

Here we report the expression profiles during seed development of over 1400 *M. truncatula* genes encoding known and putative transcription factors. These genes have been classified into functional groups. A hierarchical clustering was performed, thus allowing the identification of distinct profiles. Genes expressed during different phases of seed development were identified. We are focusing on genes encoding seed-specific transcription factors preferentially expressed just before or during the massive accumulation of storage reserves. These factors may represent regulators able to control the expression of genes ensuring storage functions. Once these genes have been identified, mutant alleles will be screened within a TILLING resource for *M. truncatula* established in our laboratory.

#### Acknowledgements:

This research was supported by the FP6 EU project "Grain Legumes" (FOOD-CT-2004-506223)



### Use of *Medicago truncatula* to investigate interactions between maternal and filial tissues during legume seed development

Karine Gallardo,<sup>a</sup>, Jérôme Verdier,<sup>a</sup>, Christian Firnhaber,<sup>b</sup> H el ene Zuber,<sup>a</sup> Klementina Kakar,<sup>d</sup> Delphine Hericher,<sup>a</sup> Maya Belghazi,<sup>c</sup> Helge K uster,<sup>b</sup> Michael Udvardi<sup>d</sup> and Richard Thompson<sup>a</sup>

<sup>a</sup> UMR en G n tique et Ecophysiologie des L gumineuses   Graines (INRA-UMRLEG), Domaine d'Epoisses, 21110 Breten eres, France, <sup>b</sup> Genomics of Legume Plants, Center for Biotechnology, Bielefeld University, D-33595 Bielefeld, Germany

<sup>c</sup> INRA UMR 6175, Service de Spectrom trie de Masse pour la prot omique, 37380 Nouzilly, France, <sup>d</sup> MPI for Molecular Plant Physiology, Am M hlenberg 1, D-14476 Golm, Germany. (thompson@epoisses.inra.fr)

The kinetics of accumulation of the major seed storage products has been characterized at the gene product level in several major crop legumes, and a number of putative regulatory components have now been identified including transcription factors, nutrients and plant hormones. To deepen our knowledge of the processes determining protein yield and composition, we are exploiting the genomics resources of the model legume *Medicago truncatula*. The legume oligoarray Mt16kOLI1, has been used to dissect the networks of gene expression which unfold during seed development. This study was extended and compared to protein changes occurring during this process, resulting in a seed proteome reference map of more than 200 proteins. The individual expression patterns of the genes and proteins identified were also examined in the isolated seed coat, endosperm and embryo. A clear repartition of activities between these compartments was observed, which demonstrated the major contribution made to the embryo by the surrounding tissues, and suggested metabolites possibly implicated in transfer between compartments.

In a parallel study, the expression profiles during seed development of more than 700 transcription factors were established. By comparing these datasets, candidate transcription factors putatively active during the grain filling stage were identified along with possible targets.



## Annexe 1

Kakar, Klementina [1], Wandrey, Maren [2], Cheung, Foo [3], Xiao, Yongli [3], Wu, Hank [3], Town, Christopher [3], Schlereth, Armin [2], Czechowski, Tomasz [2], Stitt, Mark [2], Verdier, Jerome [4], Thompson, Richard [4], Torres-Jerez, Ivone [5], Udvardi, Michael K. [5].

Genome-wide analysis of transcription factor gene expression by high-throughput qRT-PCR in *Medicago*.

Transcription factors (TFs) are master control proteins in all living cells. They control transcription and influence many biological processes including cell cycle progression, metabolism, growth, development, and responses to the environment. To characterize the full complement of transcription factors encoded by the genome of the model legume *Medicago truncatula*, we used a comprehensive list of TF protein domains to query protein-coding genes annotated by the International Medicago Genome Annotation Group (IMGAG). Predicted TFs were classified into known plant families and several potentially novel TF families, based on their characteristic protein domains. TF genes are often expressed at levels too low for their transcripts to be measured by hybridization-based techniques, such as microarrays. To overcome this problem, we developed a resource to measure transcript levels of all *Medicago* TF genes (1084 so far) by quantitative RT-PCR, using gene-specific PCR primers. This has enabled us to identify many organ-specific TFs in *Medicago*. Quantitative RT-PCR analysis of gene expression at six stages of seed development, combined with hierarchical cluster analysis (HCA) facilitated the identification of numerous seed-specific TF genes, some of which may regulate seed storage metabolism and nutrient composition. Functional characterisation of selected TFs is being pursued using RNAi and other reverse-genetic approaches in *Medicago*. Additionally, orthologues of these genes are being characterised functionally in *Arabidopsis*.

Acknowledgements: This research was supported by the EU FP6 integrated project “Grain Legumes” (FOOD-CT-2004-506223), the Max Planck Society, and The Samuel Roberts Noble Foundation.

---

1 - Max Planck Institute for Molecular Plant Physiology, Metabolic Networks Department, Am Mühlenberg 1, Potsdam-Golm, 14476, Germany

2 - Max Planck Institute for Molecular Plant Physiology, Metabolic Networks Department

3 - The Institute for Genomic Research (TIGR), Plant Genomics Department

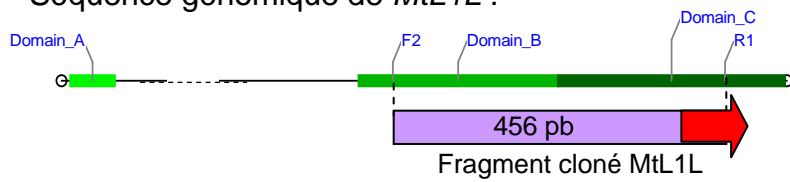
4 - Génétique et écophysiologie des légumineuses INRA Domaine d'Épouisse, Department of Genetics and Plant Breeding

5 - The Samuel Roberts Noble Foundation, Plant Biology Division

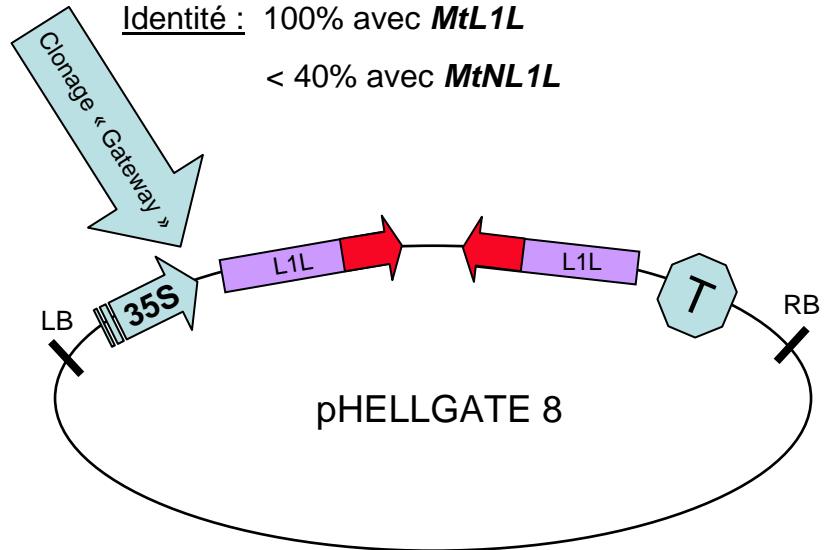


(a)

Séquence génomique de *MtL1L* :

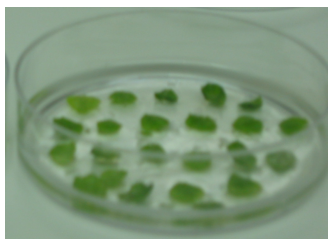


Identité : 100% avec *MtL1L*  
< 40% avec *MtNL1L*



(b)

Etapes de la transformation/régénération de *M. truncatula* R108 :



Co-culture :  
Agrobactéries -  
explants *M. truncatula*



Cals



Embyons somatiques



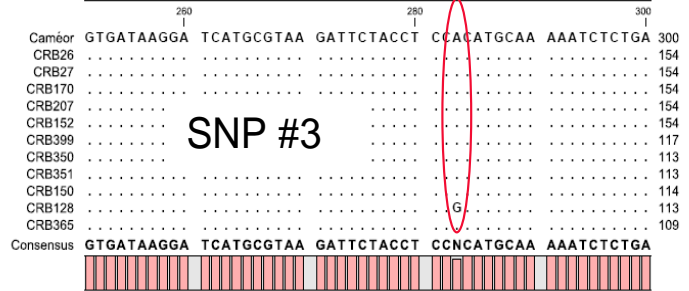
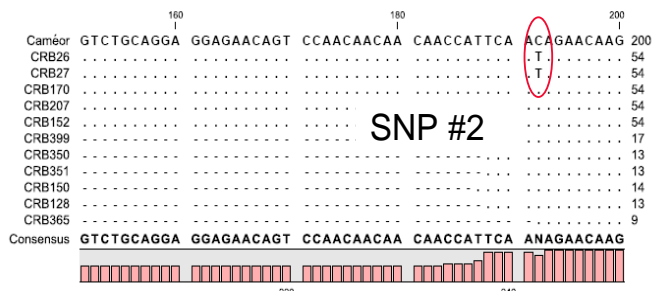
Plantes transformées

**Annexe 2 : Transformation RNAi**

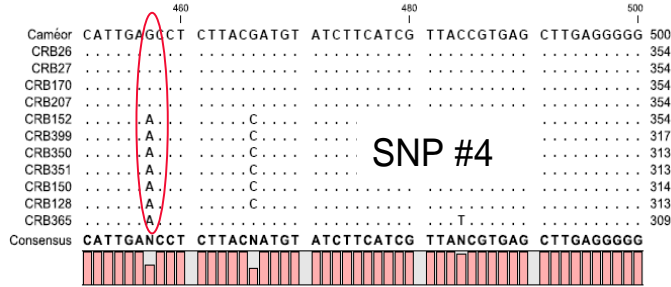
- (a) Fragment cloné et vecteur de transformation
- (b) Différentes étapes de la transformation et régénération des plantes transformées



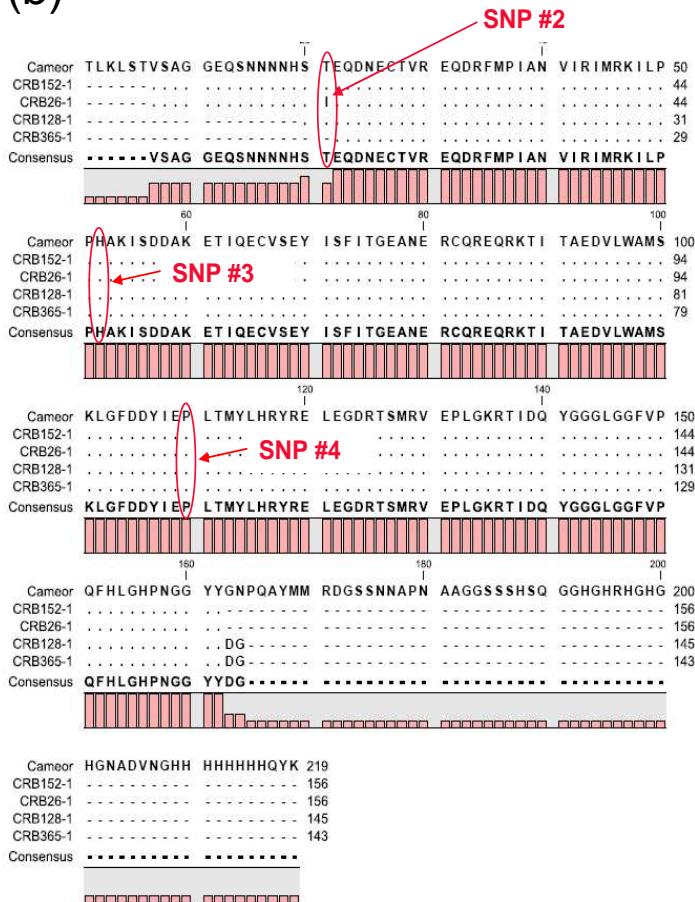




(a)



(b)



**Annexe 3 : Alignement de séquence de certains écotypes portant les différents SNPs avec l'écotype de référence, Caméor (en haut)**

- (a) Alignement des séquences nucléotidiques obtenues après séquençage d'une partie des écotypes identifiés comme polymorphe pour *PsLIL*
- (b) Alignement des séquences protéiques déduites de ces analyses de séquence





## **Contrôle transcriptionnel du développement de la graine des légumineuses**

### **Caractérisation fonctionnelle du gène *LEC1-LIKE***

La qualité nutritionnelle des graines des légumineuses constitue un enjeu économique important face aux besoins croissants en protéines végétales destinées à l'alimentation animale et humaine. Cependant, certains caractères, tels que la teneur en protéines des graines, peuvent être améliorés. Les facteurs de transcription (TFs), qui sont une composante majeure de la régulation de l'expression des gènes, interviennent dans le contrôle transcriptionnel de l'accumulation des protéines de réserve des graines. Ce travail a tout d'abord consisté (i) à définir un cadre d'étude du développement de la graine, (ii) à identifier des facteurs de transcription potentiellement impliqués dans la phase de remplissage, (iii) à ensuite valider leur rôle chez la légumineuse modèle *Medicago truncatula* et enfin leur influence sur la qualité de la graine de pois (*Pisum sativum*).

Dans un premier temps, une vision globale de l'expression de 51k gènes pendant le développement de la graine a permis de mettre en évidence les stades clefs du développement de la graine et les processus moléculaires associés. Sept stades de développement ont ainsi été choisis pour suivre l'expression de 712 facteurs de transcription (TFs) par qRT-PCR. Les 170 TFs montrant une expression différentielle pendant le développement de la graine ont été répartis dans des groupes associés à l'embryogénèse tardive, à la transition embryogénèse-remplissage, à la synthèse des vicilines, des légumines A, des légumines K et enfin à la dessiccation. Neuf facteurs de transcription candidats potentiellement impliqués dans la phase de remplissage de la graine ont été sélectionnés. L'isolement de mutants pour ces gènes a été réalisé pour valider fonctionnellement leur rôle.

Dans un deuxième temps, le rôle du facteur de transcription *LEAFY COTYLEDON1-LIKE (L1L)* a été étudié chez *M. truncatula* et *P. sativum*. Les résultats obtenus ont montré que, chez ces espèces, le gène *L1L* s'exprimait préférentiellement dans la graine au stade de déclenchement de la synthèse des protéines de réserve. L'analyse des mutants EMS et des transformants RNAi indique un rôle négatif de ce gène dans l'accumulation des protéines de réserve. Les résultats obtenus et ceux publiés chez *A. thaliana* suggèrent que *L1L* pourrait jouer un rôle antagoniste à celui de *LEC1* dans la régulation du processus d'accumulation des protéines de réserve et ainsi influencer la teneur en protéines des graines. Un polymorphisme de séquence du gène *PsL1L* corrélé à un caractère d'intérêt agronomique, le poids de 1000 grains, a également été mis en évidence confortant notre hypothèse que *L1L* est un gène fondamental dans le développement de la graine et particulièrement dans l'élaboration de la qualité de la graine chez les légumineuses.

### **Transcriptional regulation of legume seed development.**

#### **Functional analysis of the *LEC1-LIKE* gene.**

The nutritional quality of legume seeds is an important economic issue in the light of the growing demand for vegetable protein for livestock consumption and humans. Some traits involved, such as seed protein content are targets for improvement. Certain transcription factors (TFs) play an important role in the transcriptional regulation of different seed storage proteins. The goal of this study has been (i) to define a framework of seed development, (ii) to identify transcription factors potentially involved in the filling phase, (iii) and to validate their role with the model legume *Medicago truncatula* and evaluate their influence on the pea seed quality (*Pisum sativum*).

Firstly, an overview of the expression of 51k genes during seed development allowed us to identify key stages of seed development and corresponding molecular processes. Seven key stages of seed development have been chosen for the profiling of 712 transcription factors (TFs) using qRT-PCR. Out of these, 170 TFs showed differential expression during seed development. These have been divided into functional groups associated with late embryogenesis, the transition between embryogenesis and seed filling, the synthesis of vicilin, legumin A, legumin K and finally the desiccation processes. Nine candidate TFs potentially involved in the seed filling phase have been chosen and mutants have been identified for these genes, in order to validate their functional role.

Secondly, we focused on the role of the transcription factor, *LEAFY COTYLEDON1-LIKE (L1L)* in legumes. Our results showed that in *M. truncatula* and *P. sativum*, *L1L* genes were expressed preferentially at the onset of seed storage protein synthesis. Mutant analyses in both species indicated a negative role of this gene in seed storage protein accumulation of the mature seed. According to our results in *M. truncatula* and from studies in *A. thaliana* on the role of *LEC1* gene during seed development, *L1L* could play an antagonistic role to that of *LEC1* in regulating storage proteins accumulation, and finally influencing seed protein content. A sequence polymorphism of *PsL1L* correlated with the agronomic trait of the 1000 grain weight, suggesting the fundamental role of the gene *L1L* in determining legume seed quality.