



HAL
open science

Analyse structurale et transcriptionnelle des rétrotransposons du génome de la vigne, *Vitis vinifera* L.

Cédric Moisy

► **To cite this version:**

Cédric Moisy. Analyse structurale et transcriptionnelle des rétrotransposons du génome de la vigne, *Vitis vinifera* L.. Sciences du Vivant [q-bio]. Université Louis Pasteur (Strasbourg 1), 2008. Français. NNT: . tel-02820878

HAL Id: tel-02820878

<https://hal.inrae.fr/tel-02820878>

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.

Thèse de doctorat de l'Université Louis Pasteur de Strasbourg

Ecole doctorale des Sciences de la Vie et de la Santé
Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

Présentée par Cédric MOISY

Pour obtenir le titre de Docteur de l'Université Louis Pasteur de Strasbourg

Sujet de la thèse :

**Analyse structurale et transcriptionnelle des
rétrotransposons du génome de la vigne, *Vitis vinifera* L.**

Soutenue le 19 mars 2008, à l'INRA de Colmar

Devant le jury composé de :

Serge	Potier	Professeur	Président du jury
Serge	Delrot	Professeur	Rapporteur
Olivier	Panaud	Professeur	Rapporteur
Elisabeth	Huguet	Maître de Conférence	Examineur
Frédérique	Pelsy	Chargée de Recherche	Directeur de thèse

REMERCIEMENTS

Je tiens à remercier tous les membres de mon jury d'avoir accepté d'évaluer mon travail.

Merci à Serge Delrot d'avoir tout de suite accepté de faire partie de mon jury malgré un emploi du temps que j'imagine très chargé.

Merci à Serge Potier de m'avoir suivi pendant ces trois années et d'avoir participé activement aux réunions de mon comité de suivi de thèse.

Merci à Olivier Panaud pour ta disponibilité et pour les cours passionnants que tu as dispensés lorsque j'étais en DESS, car ils ont grandement contribué à ma décision de faire une thèse sur cette thématique.

Merci à Elisabeth Huguet d'avoir guidé, il y a quelques années déjà, mes premiers pas dans un laboratoire et accompagné mes premières PCR. Merci pour ton soutien et pour les conseils que tu m'as donnés depuis.

- - - - -

Laboratoire Génétique et Amélioration de la Vigne (INRA, Colmar)

Merci du fond du cœur à Frédérique pour tout ce que tu m'as apporté, pour ton enthousiasme, ta patience, ta disponibilité et toutes tes autres qualités humaines et professionnelles. On ne peut pas rêver meilleur encadrement !

Mille mercis à Emilce de m'avoir accueilli comme un membre de sa famille lors de mon arrivée à l'INRA, pour ta bonne humeur, tes éclats de rire communicatifs et ton sourire. Merci aussi pour ton soutien inébranlable pendant ces trois années, pour tous les bons petits plats que j'ai pu savourer chez toi, et cet inoubliable poulet à *la Gaston Gérard* ! Merci aussi à ta famille et tes joyeux enfants Rodolphe et Lara, qui m'ont accueilli comme un grand frère. Merci de m'avoir fait confiance pour m'occuper d'eux car nous avons partagé des moments inoubliables.

Merci à Pascale pour ton soutien, pour ton aide spontanée lors des longues séances de bouturages et pour m'avoir fait découvrir les joies des inoculations de mildiou !

Merci à Vincent pour ton aide précieuse, pour tes conseils, tant sur la vigne qu'en informatique, mais aussi pour ton sens de l'humour !

Merci à Pere, Charles et Eric d'avoir toujours pris le temps de répondre à mes interrogations lors de mes fréquentes apparitions dans vos bureaux, pour votre enthousiasme et pour les discussions enrichissantes que j'ai eues avec vous.

Merci à tous les stagiaires qui ont participé directement ou indirectement à ma thèse et que j'ai eu le plaisir d'encadrer, tout particulièrement à Cindy et Marie.

Merci à Gigi d'avoir accompagné mes premiers pas au labo.

Merci à Lucie d'avoir participé à ce travail.

Merci à tous les autres membres de l'équipe Génétique et Amélioration de la Vigne d'avoir partagé mon quotidien (et de nombreux pots !) depuis mon premier stage à l'INRA et pour m'avoir fait découvrir les plaisirs de la dégustation du vin.

Merci à Didier de m'avoir permis de travailler dans cette équipe.

- - - - -

Je tiens également à remercier Hélène Lucas et Alexandre de Kochko d'avoir participé activement aux réunions de mon comité de suivi de thèse, pour vos conseils et les discussions intéressantes que nous avons eues à propos de mon travail.

Merci à Xavier Daire et à Adrien Gauthier de l'UMR Plante Microbe et Environnement de l'INRA de Dijon d'avoir guidé mes premiers Northern et, de ce fait, accepté de partager avec moi quelques rayonnements radioactifs !

Merci à Keith Garisson pour sa collaboration à la détection des nouvelles familles de rétrotransposons.

- - - - -

Je tiens aussi à remercier tous les membres de l'équipe Virologie qui m'ont toujours accueilli à bras ouverts et aussi beaucoup aidé.

Merci à Gérard pour ton soutien et ton aide dans la dernière ligne droite, ainsi que pour tous les bons moments partagés en dehors du labo.

Merci à Laure pour tes encouragements et pour avoir spontanément proposé de relire mon manuscrit.

- - - - -

Merci à Jacky et à toutes les personnes qui m'ont accompagné dans la mise en place de mes populations de plantes, et qui se sont intéressées à mon projet.

Merci à l'équipe du Laboratoire de Biologie du Développement de la Vigne de m'avoir fourni les échantillons d'embryons somatiques. Un merci particulier à Claude d'avoir toujours répondu à mes besoins urgents en azote liquide. Sans toi, beaucoup de prélèvements seraient partis à la poubelle !

Merci également à tout le personnel du centre INRA de Colmar avec qui j'ai partagé de bons moments, autour d'un repas, d'un café ou d'un barbecue.

Merci à Sabine pour ta gentillesse au quotidien.

Merci à toutes les personnes qui, parce qu'elles m'ont offert un sourire ou un « bonjour » lorsque je les croisais dans un couloir, ont rendu le quotidien encore plus agréable.

Enfin, merci à toi, PowerBook G4, mon compagnon de route qui ne m'a jamais laissé tombé malgré les relations conflictuelles que nous avons pu entretenir... souvent à cause de tes plantages « *inopinés* », comme tu aimes les qualifier.

- - - - -

Merci à mes parents, de tout mon cœur, de m'avoir soutenu, encouragé et montré votre amour en toutes circonstances. Le simple fait de savoir que je peux toujours compter sur vous est, de toutes vos qualités, celle qui m'apporte probablement le plus.

Merci à ma P'tite Puce pour tout ce que tu m'as apporté pendant ces trois années, pour ton amour et ton soutien, et pour avoir supporté les contraintes de la vie avec un doctorant en phase terminale de rédaction.

Merci à mon frère qui est toujours là, à mes côtés. Merci aussi à Eva qui m'a fait avec toi ce merveilleux cadeau en m'offrant le bonheur d'être tonton. Et merci au plus beau des bébés, le p'tit Noam, qui m'a apporté l'un des plus beaux moments de ma vie en venant au monde. Ton arrivée à ce moment de ma thèse m'a aidé à relativiser bien des choses !

Merci à mes grands-parents pour votre soutien, même si mon domaine de recherche est un peu obscur pour vous.

Merci à mes amis d'enfance : Mathieu, Tito, Vince et Maxou. Ce n'est pas toujours facile de comprendre le métier des autres... mais qu'est-ce qu'on est bien quand on est tous ensemble ! Merci aussi à tous les autres membres de la bande, dispersés aux quatre coins de la France : Vincent, Julien, Mimie, Cindy et Mickaël, Sabine et JB, Virginie, Charlotte et tous les autres... pour les bons moments passés et à venir.

Merci aux alsaciens « pur souche », de belles rencontres ! A Matthieu de m'avoir suivi et bien chambré lors de mes périodes de bénévolat forcé, pour nos débats enflammés et les soirées inoubliables que nous avons passées ensemble. A Sylvain dont le chemin suit le mien depuis la première extraction d'ADN qu'on a réalisée ensemble, de m'avoir fait découvrir l'Alsace et ses merveilles. Au jeune marié J-M pour nous avoir souvent accueilli chez lui. Merci à tous les trois pour ces randos VTT aussi épuisantes que dépaysantes. Un gros bisou à vos chéries Céline, Françoise et Sophie.

- - - - -

Merci à l'ADAS-INRA pour tous les bons moments que j'ai passés lors des sorties collectives, autour d'un feu de camp ou lors d'une nuit à la belle étoile dans les Vosges, et de m'avoir permis de découvrir de nouveaux sports et de faire autant de belles rencontres.

Merci à tous mes amis sportifs, avec qui je me suis défoulé en sortant des manip', pour tous les moments intenses que nous avons passés ensemble.

Enfin, merci à aux membres de mon équipe de tennis de table de choc, j'ai nommé Pere, Gérard, Gisèle et Gillou, et à ceux des autres équipes de l'INRA, tout particulièrement à Manu, Pascale, Sabine et Guillaume, ainsi qu'à mes coéquipiers de Jorkyball pour ces matches enflammés... et les petites mousses qui ont suivies !

Sommaire

SOMMAIRE

INTRODUCTION GÉNÉRALE	1
PARTIE I : INTRODUCTION.....	4
CHAPITRE I : LA VIGNE.....	5
<i>I.1 - Présentation de l'espèce Vitis vinifera L.....</i>	<i>5</i>
I.1.1 - Taxonomie	5
I.1.2 - Morphologie de la vigne.....	6
a) L'appareil végétatif.....	6
b) L'appareil reproducteur	6
I.1.3 - Physiologie de la vigne.....	7
a) Cycle de développement végétatif.....	7
b) Cycle de développement reproducteur	8
c) Conduite de la vigne	8
I.1.4 - La filière viticole en France	9
I.1.5 - Érosion de la diversité variétale	10
I.1.6 - Le génome de la vigne.....	11
<i>I.2 - Constitution des cépages et variétés actuels</i>	<i>12</i>
I.2.1 - Notions de cépage, de variété et de clone.....	12
a) Le cépage.....	12
b) La variété.....	12
c) Le clone	13
I.2.2 - Modes de propagation de la vigne.....	13
a) La reproduction sexuée	13
b) La multiplication végétative.....	14
c) Conséquences génétiques de la multiplication végétative	15
I.2.3 - Scénarios pour la constitution des cépages.....	15
a) Origine polyzygotique.....	15
b) Origine monozygotique	15
<i>I.3 - Origines des mutations somatiques accumulées au cours des cycles de</i> <i> multiplication végétative.....</i>	<i>17</i>
I.3.1 - Accumulation de mutations somatiques	17
I.3.2 - Origines des mutations somatiques chez les plantes à multiplication végétative	
.....	18
a) Mutations ponctuelles, insertions et délétions.....	18
b) Les changements chromosomiques	19
c) Chimères et réarrangements cellulaires	19
d) L'activité des éléments transposables.....	20
CHAPITRE II : LES ÉLÉMENTS TRANSPOSABLES	21
<i>II.1 - Evolution des pensées</i>	<i>21</i>
<i>II.2 - Classification des éléments mobiles</i>	<i>25</i>
II.2.1 - Les éléments de classe I : les rétrotransposons.....	25
a) Les rétrotransposons à LTR	25
b) Les rétrotransposons (ou rétrotransposons sans LTR).....	26
II.2.2 - Les éléments de classe II : les transposons.....	27
II.2.3 - Vers une nouvelle classification ?.....	28
II.2.4 - Autonomie et complémentarité fonctionnelle.....	29

a) Les éléments autonomes.....	29
b) Les éléments non-autonomes	29
c) Mobilité des éléments non-autonomes par complémentation fonctionnelle	29
d) Transcription et mobilité.....	31
II.2.5 - Le cycle de rétrotransposition des rétrotransposons à LTR	31
a) Transcription d'un ARN messager	31
b) Traduction	31
c) Encapsidation.....	32
d) Synthèse de la nouvelle copie ADN du rétrotransposon	33
e) Intégration dans le génome hôte.....	34
II.2.6 - Principe de l'estimation de l'âge des copies (datation).....	34
II.2.7 - La notion de « famille » chez les rétrotransposons.....	35
a) Erreurs lors de la transcription inverse	35
b) Divergence des copies.....	35
c) Evolution des domaines fonctionnels chez les rétroéléments	36
<i>II.3 - Contrôle de l'activité des éléments transposables.....</i>	<i>38</i>
II.3.1 - La sélection : élimination de copies	38
II.3.2 - Régulation de l'activité des éléments transposables	38
a) Régulation au cours du développement	39
b) Activation des éléments mobiles en conditions de stress	40
II.3.3 - Interférence ARN.....	45
<i>II.4 - Participation des éléments transposables à la structure et à la dynamique des génomes</i>	<i>46</i>
II.4.1 - Impact sur la taille du génome hôte	46
a) Invasion du génome	46
b) Recombinaison et perte de matériel génomique.....	47
II.4.2 - Participation des éléments transposables à la structure du chromosome	48
a) Les éléments transposables comme constituants des centromères	48
b) Participation au maintien des télomères	48
II.4.3 - Influence sur l'expression des gènes de l'hôte	49
a) Insertion d'éléments transposables dans les gènes	49
b) Régulation épigénétique de l'expression de gènes de l'hôte	49
<i>II.5 - Les rétrotransposons du génome de la vigne</i>	<i>53</i>
II.5.1 - L'élément Vine-1	53
II.5.2 - La famille Tv1.....	54
II.5.3 - L'élément Gret1	55
II.5.4 - Les éléments transposables du génome de PN40024.....	55
CHAPITRE III : PRÉSENTATION DES TRAVAUX DE THÈSE.....	57
PARTIE II : LES RÉTROTRANSPOSONS DU GÉNOME DE LA VIGNE	59
CHAPITRE I : CARACTÉRISATION DE NOUVELLES FAMILLES DE RÉTROTRANSPOSONS À LTR 60	
I.1 - Méthode de Pearce et al.	60
I.2 - Distinction de 10 nouvelles familles de rétrotransposons à LTR.....	60
<i>Characterisation of ten novel Ty1 copia-like retrotransposon families of the grapevine genome</i>	<i>63</i>
CHAPITRE II : LES RÉTROTRANSPOSONS « V ».....	79
II.1 - Des rétrotransposons de vigne annotés dans les banques de données.....	79
II.2 - Prise en compte de 5 familles V dans notre étude.....	79
a) Recherche de la séquence codante des copies V1, V11, V12, V13a, V13b et V14	80

b) Evaluation du nombre d'éléments « V » dans le génome de PN40024.....	80
CHAPITRE III : CONCLUSION	82
PARTIE III : ÉTUDE DE L'ACTIVITÉ TRANSCRIPTIONNELLE DES	
RÉTROTRANSPOSONS.....	84
CHAPITRE I : RECHERCHE DE RÉTROTRANSPOSONS DANS LES BANQUES EST DE VIGNE	85
I.1 - Blast sur les banques de séquences EST de vigne	85
I.2 - Choix des séquences requêtes.....	86
I.3 - Résultats	86
a) Représentativité des 18 familles.....	86
b) EST détectées dans plusieurs variétés	86
c) EST de rétrotransposons plus nombreuses en conditions de stress	87
d) Position des EST sur les séquences requêtes	88
I.4 - Discussion.....	89
a) Transcription des rétrotransposons chez la vigne.....	89
b) Un niveau d'expression plus fort que chez d'autres plantes ?	91
c) Conditions d'expression.....	91
d) Représentativité des 18 familles.....	92
e) Lien entre la taille des séquences requêtes et le nombre d'EST détectées ?.....	94
f) Position des EST et séquences chimériques	94
I.5 - Conclusion	96
CHAPITRE II : ÉVALUATION DU NIVEAU DE TRANSCRIPTION PAR RT-PCR SEMI-	
QUANTITATIVE	97
II.1 - Avant propos.....	97
II.2 - Dispositif expérimental	98
II.2.1 - Le matériel végétal.....	98
a) Choix de la variété Riesling	98
b) Mise en place de collections de boutures.....	98
c) Prélèvement des échantillons d'organes	99
d) Embryons somatiques	100
II.2.2 - Stress biotique et abiotique.....	100
a) Préparation des plantes.....	100
b) Application des stress	101
II.2.3 - Extraction des ARN totaux.....	102
a) Extraction	102
b) Elimination de l'ADN génomique résiduel.....	102
c) Vérification de la qualité des ARN extraits.....	103
II.2.4 - Synthèse des ADN complémentaires (ADNc)	103
a) Etalonnage des quantités d'ARN.....	103
b) Synthèse des ADNc simple-brin	104
II.2.5 - Témoin interne : le gène de l'actine.....	104
a) Test PCR sur ARN totaux	104
b) Evaluation de l'intégrité des ADNc.....	105
II.2.6 - Evaluation de la quantité relative de transcrits par PCR semi-quantitative ...	105
a) Dessin des amorces spécifiques des rétrotransposons	105
b) Gènes témoins de l'effet des stress.....	106
c) Etalonnage des quantités d'ADNc	106
II.3 - Résultats	107
II.3.1 - Détermination du niveau de transcription dans les organes	107
II.3.2 - Evaluation de l'influence de deux stress sur le niveau de.....	108

II.3.2 - transcription dans les feuilles.....	108
a) Blessure mécanique	108
b) Inoculation par le mildiou	109
II.3.3 - Etude de la variabilité des séquences transcrites	111
a) Variabilité de la région gag des transcrits	111
b) Variabilité de la région UTL des transcrits TvV1	112
<i>II.4 - Discussion des résultats de l'analyse du niveau de transcription par RT-PCR ...</i>	<i>113</i>
a) Transcription des rétrotransposons dans les organes	113
b) Influence des stress sur le niveau de transcription dans les feuilles.....	115
CHAPITRE III : EVALUATION DE LA TAILLE DES TRANSCRITS	119
<i>III.1 - Détection des transcrits de 4 rétrotransposons par Northern</i>	<i>119</i>
<i>III.2 - Résultats</i>	<i>120</i>
<i>III.3 - Conclusions</i>	<i>120</i>
CHAPITRE IV : DISCUSSION DE L'ÉTUDE DU NIVEAU DE TRANSCRIPTION DES RÉTROTRANSPOSONS CHEZ LA VIGNE.....	122
PARTIE IV : EVOLUTION DE LA FAMILLE TVV1	126
CHAPITRE I : ÉTUDE DE LA VARIABILITÉ STRUCTURALE DES ÉLÉMENTS TVV1	127
CHAPITRE II : CARACTÉRISATION D'UN RÉTROTRANSPOSON DU GÉNOME DE <i>S. DEMISSUM</i> PRÉSENTANT DES SIMILARITÉS AVEC TVV1	141
II.1.1 - Résultats	141
a) Description de l'élément TvV1-SD	141
b) Relations phylogénétiques avec les autres rétrotransposons de plantes	142
c) Comparaison de l'élément de <i>S. demissum</i> à ceux de la famille TvV1 de vigne	143
II.1.2 - Conclusions.....	144
PARTIE V : MATÉRIELS ET MÉTHODES.....	147
<i>Obtention de boutures</i>	<i>148</i>
<i>Extraction des ARN totaux</i>	<i>149</i>
<i>Northern blot</i>	<i>150</i>
CONCLUSION GÉNÉRALE.....	152
BIBLIOGRAPHIE	158
ANNEXE	172

Introduction générale

Introduction générale

La vigne *Vitis vinifera* L. est cultivée depuis l'Antiquité et comprend plus de 6000 variétés, dont les plus anciennes seraient âgées de centaines, voire de milliers d'années. La multiplication végétative, qui fait partie du processus de domestication de nombreuses espèces, a permis très tôt de multiplier des génotypes sélectionnés pour leurs caractères d'intérêts agronomiques.

Les variétés de vigne actuelles sont composées d'un ensemble de clones présentant des caractéristiques ampélographiques homogènes. Ces clones sont tous issus de la multiplication végétative d'un individu unique afin d'assurer le maintien de l'identité génétique entre la plante initiale et les plantes filles. Cependant, des mutations spontanées peuvent apparaître, s'accumuler au cours des cycles de multiplication végétative et être, dans certains cas, à l'origine de nouveaux phénotypes présentant des aptitudes culturales et technologiques différentes de la plante initiale. En l'absence de reproduction sexuée, l'accumulation de mutations somatiques représente l'unique source de variation génétique chez les plantes. Plusieurs mécanismes peuvent générer de la variabilité génétique au sein d'un individu multiplié végétativement, notamment les éléments mobiles présents dans les génomes.

Les éléments transposables sont des éléments génétiques mobiles, capables de se multiplier dans le génome, parfois de manière autonome. Ces caractéristiques leur ont longtemps valu le qualificatif « *d'ADN égoïste* ». Découverts dans les années 1940, ils ont depuis été trouvés dans tous les génomes dont ils sont parfois les constituants majeurs. Ils représentent par exemple 45% du génome humain et plus de 60% du génome du maïs. Leur nombre et leur caractère invasif font de ces éléments des acteurs majeurs de la structure et de l'expression des génomes et ils sont désormais considérés comme une source de diversification et d'évolution qui participe à l'adaptation des génomes en réponse aux changements environnementaux.

L'induction de leur transcription, qui constitue la première étape du cycle conduisant à l'insertion d'une nouvelle copie dans le génome, a été mise en évidence en réponse à différents stress biotiques et abiotiques. Or la vigne est une plante pérenne qui subit de

nombreux stress durant sa longue période de production, pouvant conduire à l'activation des éléments transposables présents dans son génome et à l'apparition de nouvelles insertions.

Mon projet de thèse s'inscrit dans un contexte général de compréhension des mécanismes biologiques qui participent à l'évolution des génomes au cours des cycles de multiplication végétative. Au commencement de mon travail, seulement trois rétrotransposons du génome de la vigne étaient décrits : la famille *Tvv1* et les deux copies uniques *Vine-1* et *Gret1*. Mon projet a évolué au cours de son développement, en particulier suite à la publication de la séquence du génome de la vigne, au milieu de l'année 2007, et peut se résumer à une étude de la structure et de la dynamique d'expression des rétrotransposons du génome de la vigne.

Mon manuscrit de thèse est divisé en quatre parties :

La première partie est une synthèse bibliographique qui a pour but d'introduire le contexte scientifique dans lequel ce travail a été réalisé. Cette introduction est subdivisée en trois chapitres. Le chapitre 1 présente la vigne, son mode de propagation et ses conséquences sur le génome et sur la constitution des variétés. Le chapitre 2 débute par l'histoire de l'étude des éléments mobiles, de leur découverte en 1940 à nos jours. Il se poursuit ensuite par la description des éléments transposables, plus particulièrement la classe des rétrotransposons à LTR, la régulation de leur activité et leur impact sur la dynamique et l'expression du génome. Enfin, le chapitre 3 présente les questions sur lesquelles mon travail de thèse tente d'apporter un éclairage.

Les quatre parties suivantes sont consacrées à la présentation des principaux résultats que j'ai obtenus pendant ma thèse ainsi qu'aux matériels et les méthodes utilisés au cours de ces travaux.

Enfin, le manuscrit se termine par une discussion générale qui souligne les conclusions importantes de mon travail et qui présente les perspectives envisageables.

En annexe est présenté un article de revue que j'ai rédigé au début de ma thèse pour la revue *Biofutur* (Elsevier) sur la régulation des éléments transposables.

Partie I : Introduction

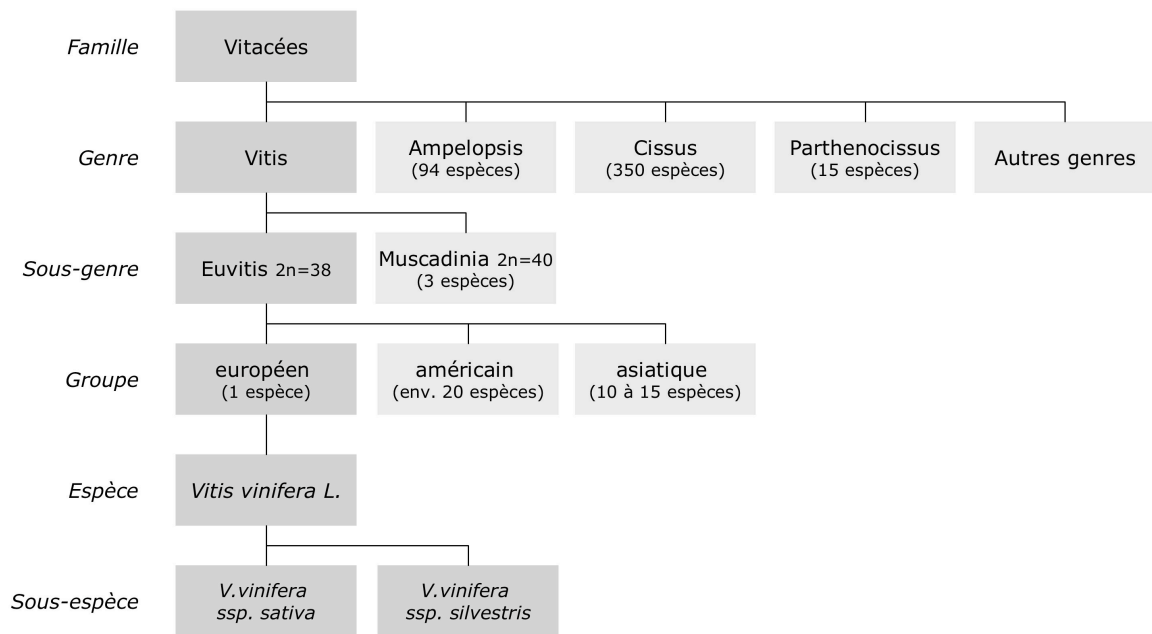


Figure 1 : Eléments de systématique des Vitacées, d'après Galet (1993)
(selon la classification considérée, *Muscadinia* peut être considéré comme un sous-genre de *Vitis* ou comme un genre à part entière).

Chapitre I : La vigne

I.1 - Présentation de l'espèce *Vitis vinifera* L.

I.1.1 - Taxonomie

La vigne est une plante diploïde appartenant à la famille des Vitacées. Parmi les 18 genres qui composent cette famille, le genre *Vitis* regroupe la plupart des espèces pérennes et cultivées dont les fruits sont consommables. Ce genre est séparé en deux sections, *Muscadinia* et *Euvitis*, qui se distinguent par quelques caractères morphologiques et anatomiques, ainsi que par leur nombre de chromosomes (Figure 1).

La section *Muscadinia* comprend trois espèces diploïdes ($2n = 40$ chromosomes), originaires des régions chaudes et tempérées d'Amérique du Nord, notamment l'espèce *rotundifolia* qui est utilisée dans des programmes de sélection pour sa résistance ou tolérance à de nombreux pathogènes de la vigne, comme le mildiou, l'oïdium et le phylloxéra.

La section des *Euvitis* regroupe les vignes dites « vraies » et comporte près de soixante espèces diploïdes ($2n = 38$ chromosomes). Durant l'histoire de la vigne, les périodes de glaciation successives ont provoqué la séparation de trois refuges historiques à partir desquels la plante s'est ensuite dispersée. Il en résulte aujourd'hui trois groupes géographiques distincts : les vignes américaines, qui sont composées d'une vingtaine d'espèces, dont certaines ont été utilisées pour la sélection de porte-greffes résistants au phylloxéra (*V. rupestris*, *V. riparia*...); les vignes asiatiques, qui sont représentées par une quinzaine d'espèces, dont l'espèce *V. amurensis* qui est utilisée dans des programmes de sélection par hybridation inter-spécifique pour sa résistance au froid ; et la vigne européenne, qui n'est, quant à elle, constituée que d'une seule espèce, *V. vinifera* L., qui regroupe l'ensemble des cépages cultivés dans nos régions, ainsi que les vignes sauvages de type dioïque, *V. vinifera* ssp. *silvestris*. Le compartiment cultivé de l'espèce, *V. vinifera* ssp. *sativa*, comprend plus de 6000 variétés ou cépages, qui présentent des caractéristiques morphologiques ou ampélographiques très différentes (Alleweldt and Possingham, 1988). Ce nombre est en perpétuelle évolution du fait de la disparition de certains cépages et de la création de nouvelles variétés.

1.1.2 - Morphologie de la vigne

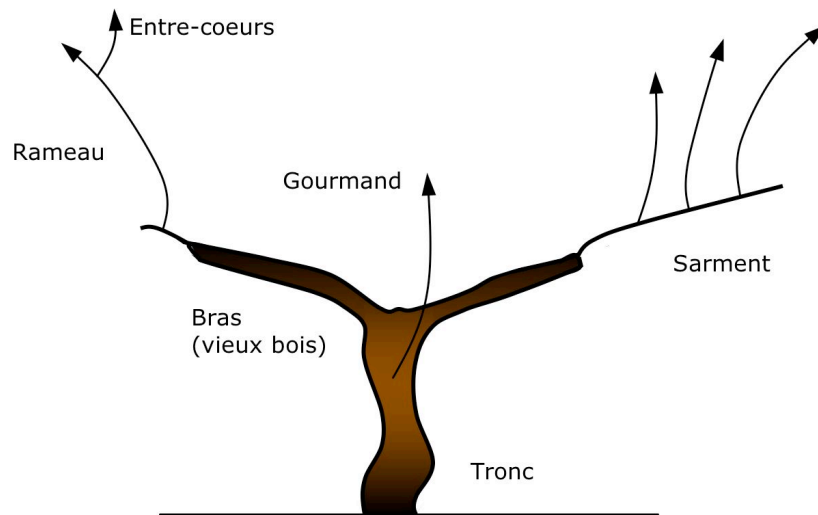
a) L'appareil végétatif

L'appareil végétatif de la vigne est formé d'organes pérennes : le tronc, les sarments et les racines ; et d'organes annuels : les feuilles, les fleurs et les fruits.

Le tronc, également appelé le « vieux bois », est généralement de forme sinueuse et recouvert d'une écorce rugueuse et filamenteuse. Au cours des années, il se ramifie pour donner un rameau, puis des sarments (Figure 2). Le rameau se développe chaque printemps à partir d'un bourgeon situé soit sur le tronc, soit sur un sarment. Chaque année après l'aoûtement, pendant lequel l'écorce fait son apparition, le rameau passe à l'état de sarment et prend une teinte brune. Les bourgeons naissent à l'aisselle d'une feuille et sont constitués d'écailles marron, imbriquées les unes dans les autres. Il existe deux types de bourgeons : le « prompt bourgeon » qui donne naissance à des entre cœurs et qui n'est pas fertile ; et le « bourgeon latent » qui est fertile et donne des inflorescences au printemps, après une période de dormance. Les feuilles sont insérées sur les rameaux au niveau des nœuds de manière alterne. Elles sont généralement constituées de cinq lobes, cinq nervures principales et cinq sinus, dont un sinus pétiolaire (Figure 2). Leur forme et leur taille dépendent fortement du cépage considéré.

b) L'appareil reproducteur

Les différents organes de reproduction, l'inflorescence, la grappe et les baies, se forment à partir de l'initiation florale qui a lieu dans les bourgeons latents. Cette initiation se déroule vers le mois d'août de l'année précédant la floraison. L'inflorescence est composée d'une rafle sur laquelle les fleurs sont regroupées en bouquets. Avant la floraison, les pétales se détachent et sont éliminés. Les fleurs de la vigne cultivée sont le plus souvent hermaphrodites, mais il existe des vignes sauvages dioïques. Le nom « grappe » est réservé à l'inflorescence fécondée qui porte les baies. Celles-ci sont de formes très variées, de tailles et de couleurs diverses selon les variétés. La baie est composée de la pulpe enveloppée dans une pellicule (Figure 2). La pulpe est formée de plusieurs assises cellulaires dont les cellules s'accroissent fortement au cours de la maturation. À la maturité, les membranes de ces cellules se gélifient et la pulpe devient souvent juteuse. La pellicule est composée d'un épiderme revêtu extérieurement d'une cuticule épaisse. L'épiderme est coloré par des molécules d'anthocyanes



Cep de vigne

Le tronc se divise en deux bras (vieux bois) portant des bois de taille (sarments) sur lesquels les bourgeons latents vont se développer pour donner les rameaux. Les rameaux se ramifient ensuite en entre-cœurs. Des gourmands se forment à partir des bourgeons présents sur le vieux bois.

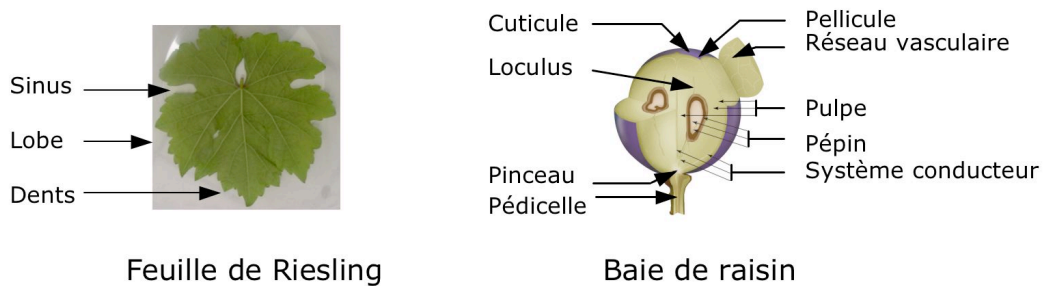


Figure 2 : Morphologie de la vigne.

(Baie d'après Carbonneau *et al.* 2007)

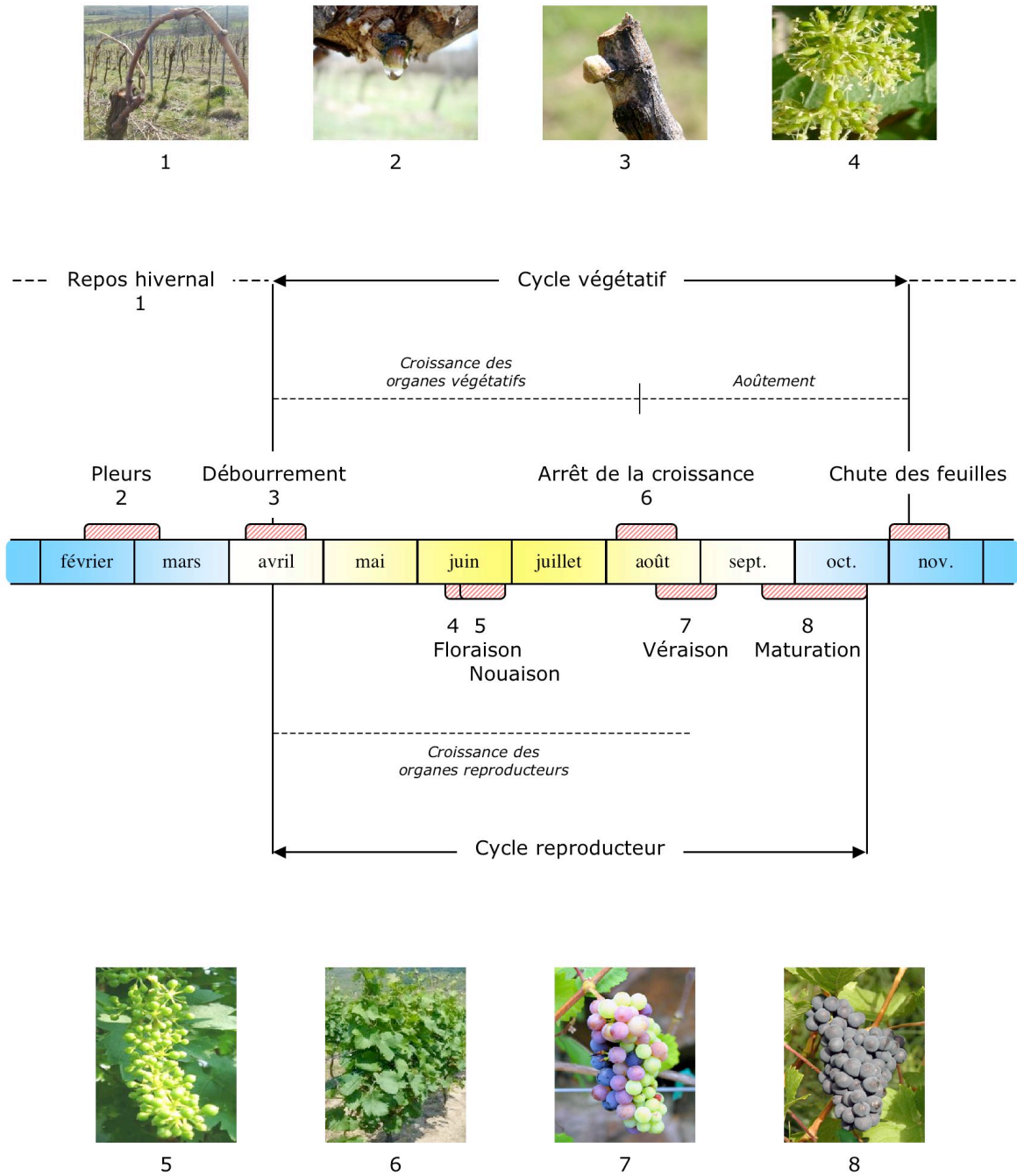


Figure 3 : Cycles de développement végétatif et reproducteur de la vigne.

contenues dans les vacuoles. La baie de raisin contient généralement entre 1 et 4 pépins qui résultent du développement d'ovules fécondés.

1.1.3 - Physiologie de la vigne

a) Cycle de développement végétatif

Il faut à la vigne un premier cycle de croissance avant de pouvoir produire des organes reproducteurs ; la jeune plante est donc stérile la première année. Le cycle végétatif est composé de six étapes (Figure 3). Les « *pleurs* » représentent le premier signe de la reprise de l'activité de la vigne après le repos hivernal. Ce phénomène se manifeste par le suintement de sève brute à la surface des plaies de tailles fraîches. Après la période des pleurs, la souche va reprendre sa « *croissance* ». La première manifestation visible de cette croissance, le « *débourrement* », est le gonflement des bourgeons jusqu'à ce que ses écailles se séparent. Cette étape correspond aussi à la sortie des premières feuilles. Au mois d'avril, une pointe verte plus ou moins saillante est visible et entraîne le rejet de la bourre (ou coton) qui a protégé l'organe pendant l'hiver. La croissance est ensuite le résultat de la multiplication et de la dilatation des cellules à partir des méristèmes. Elle se caractérise par l'allongement des rameaux et le développement des jeunes feuilles préformées dans les bourgeons. La phase de croissance s'étale depuis la fin du débourrement (avril) jusqu'à la maturation des grappes (fin de l'été). Vers la fin du cycle végétatif, vient ensuite l'étape de « *défeuillaison* » : les feuilles jaunissent, se vident des produits de la photosynthèse et tombent (novembre). Après la chute des feuilles, la vigne entre dans une période de « *repos hivernal* ». La plante ne présente plus d'activité végétative apparente et les bourgeons sont dans une phase de dormance. La fin du cycle végétatif est marquée par « *l'aoûtement* », qui se manifeste par une modification morphologique et anatomique des rameaux, notamment leur lignification. Au niveau fonctionnel, cette dernière étape s'accompagne d'une accumulation d'amidon dans la plante juste après la fin de sa croissance, et se poursuit jusqu'à la chute des feuilles. Lorsqu'une portion de sarment a dépassé le stade de bois aoûté, elle entre dans sa période de repos. Si le cycle végétatif est indispensable à la vie de la vigne, l'aoûtement est nécessaire à la pérennité de la souche ou du sarment d'une année sur l'autre.

b) Cycle de développement reproducteur

Le cycle reproducteur débute par la phase « *d'initiation florale* » durant laquelle les inflorescences se forment dans les bourgeons latents. Mais elles ne seront cependant visibles que l'année suivante, après la période de repos hivernal. Les inflorescences sortent des bourgeons au début de la période de croissance, quelques jours après le débourrement. Lorsque les feuilles sont développées, les inflorescences deviennent bien visibles, puis les boutons floraux se séparent progressivement. La « *floraison* », qui a généralement lieu vers le mois de juin (Figure 3), commence par l'ouverture de la corolle sous la pression des étamines. La corolle se détache, dessèche et tombe. L'ovaire reste alors nu, entouré par les étamines. L'ouverture des sacs polliniques se produit généralement après la chute des pétales et la « *fécondation* », rencontre d'un grain de pollen et d'un ovule, peut alors avoir lieu. La vigne étant une espèce allogame, la fécondation croisée peut se réaliser naturellement, avec néanmoins une possibilité d'autofécondation. Au cours de la « *nouaison* », l'ovaire croît pour donner le grain de raisin ou baie, alors qu'à l'intérieur de l'ovaire, les ovules fécondés vont évoluer en graines ou pépins. Les parois de l'ovaire se gélifient et la cloison qui sépare les loges ovariennes disparaît. La « *véraison* » représente le passage de la baie de l'état de grain dur, vert et acide, à celui de grain mou, coloré et souvent sucré (Figure 3). La grappe de raisin va ensuite se concentrer en sucre et en eau pour arriver à maturité vers le mois de septembre ou octobre.

La connaissance du cycle de développement de la vigne est importante pour définir les organes qui sont étudiés dans la Partie III de ce manuscrit.

c) Conduite de la vigne

Tout au long de sa vie, la vigne va subir de nombreuses interventions humaines, principalement liées à des objectifs de production. Le greffage puis la plantation des plants greffés dans la parcelle sont parmi les premiers événements qui marquent le début d'une nouvelle période de culture. La vigne étant une liane qui se développe naturellement de façon importante, la taille au cours de l'hiver permet de limiter l'allongement des sarments, de contrôler le nombre de bourgeons et donc de maîtriser la production et la vigueur des ceps. Les sarments sont ensuite rabattus (arcure, palissage). Pendant le développement végétatif, les rameaux en croissance sont à nouveau palissés et peuvent être taillés, soit sur les côtés (rognage), soit sur le dessus de la plante (écimage). Après la période des vendanges, des

engrais organiques et minéraux sont généralement apportés au sol afin d'éviter les carences lors de la prochaine période de croissance de la vigne.

En plus des interventions humaines et des blessures provoquées par le passage des machines agricoles, la vigne doit faire face aux contraintes météorologiques (gel, grêle, canicule, sécheresse...) et aux maladies provoquées par des pathogènes, comme le mildiou (*Plasmopara viticola*), l'oïdium (*Erisiphe necator*), la pourriture grise (*Botrytis cinerea*), l'ESCA (des champignons lignivores), le court-noué (Grapevine fanleaf Virus), ainsi que les attaques d'insectes comme le phylloxéra (*Dactylosphaera vitifoliae*), de chenilles, de cochenilles et de vers... Pour lutter contre ces agressions, l'homme protège la vigne par des traitements phytosanitaires qui font appel à des molécules organiques ou chimiques. Ces traitements restent souvent très nombreux, malgré l'orientation progressive de la filière viticole vers une lutte plus raisonnée.

1.1.4 - La filière viticole en France

Avec les céréales, le secteur viticole est l'un des secteurs les plus importants de la production agricole française. La vigne représente seulement 3% des terres cultivées en France, mais environ 20% de la valeur des produits agricoles français. En 2005, l'hexagone se plaçait parmi les premiers producteurs avec 19% de la production mondiale de vin (source Food and Agriculture Organisation of the United Nations). Les exportations représentaient 13,9 millions d'hectolitres, essentiellement en direction de nos voisins européens (Grande-Bretagne, Allemagne), mais aussi vers les Etats-Unis et plus récemment vers les marchés asiatiques. Les Français font également partie des plus grands consommateurs, avec une moyenne de 55 litres de vin par habitant et par an en 2004 (source Onivins).

La vigne est cultivée en France dans des régions aux conditions climatiques très différentes, au climat continental (Alsace, Champagne, Bourgogne), océanique (Bordelais, Vallée de la Loire) ou méditerranéen (Vallée du Rhône, Midi, Provence et Corse). En 2003, plus de 865300 hectares de vigne étaient répartis sur environ 120000 exploitations. A lui seul, le Languedoc-Roussillon représente un tiers du vignoble français. Les vignes cultivées en France sont composées de 50 cépages principaux. Les 10 principales variétés de cuve représentent 75% de la production viticole française : le Merlot, le Grenache, l'Ugni blanc, le

Carignan, la Syrah, le Cabernet sauvignon, le Chardonnay, le Cabernet franc, le Gamay, le Pinot noir. Le territoire français produit près de 400 vins d'appellation contrôlée et 150 vins de pays. En Alsace, les vins les plus renommés sont le Riesling, le Muscat ottonel, le Gewürztraminer, le Pinot gris, le Pinot Blanc et le Crémant, un vin effervescent.

1.1.5 - Érosion de la diversité variétale

Plusieurs événements ont conduit à la réduction de la diversité génétique des populations de vignes cultivées. A partir de la deuxième moitié du XIXe siècle, les introductions successives de différentes maladies comme l'oïdium, le mildiou et surtout le phylloxéra ont provoqué un véritable bouleversement des comportements et des pratiques viticoles en Europe. L'oïdium a été observé pour la première fois en 1847 en Angleterre, après l'importation de plants contaminés venant des Etats-Unis. Avec l'apparition du mildiou et du black-rot en 1878 et 1885, les traitements phytosanitaires se sont généralisés (Pouget, 1990). Le phylloxéra a été observé pour la première fois en France en 1868, à Nîmes et à Bordeaux. Le puceron qui en est responsable a été introduit en Europe par l'importation de variétés américaines infectées et a ensuite ravagé la quasi-totalité du vignoble français et européen en l'espace de 25 ans. La crise phylloxérique a nécessité la reconstitution de ces vignobles en généralisant le greffage des cépages traditionnels sur des porte-greffes américains résistants (voir page 14). Cette crise a entraîné une perte importante de la diversité génétique, puisque seuls les cépages les plus importants économiquement ont été replantés.

L'adoption du système des Appellations d'Origine Contrôlée (AOC) en 1935 a également participé à la diminution des ressources génétiques en limitant le choix des cépages à implanter dans la zone AOC, pour une adéquation optimale avec le terroir. En 1956, le gel a détruit de nombreux vignobles et entraîné une perte de diversité au niveau intra-variétal. Les primes d'arrachage à la fin des années 1970 ont entraîné la disparition des vignobles marginaux. La sélection clonale a également joué un rôle important dans la réduction de la diversité à l'intérieur des cépages (Figure 4). Cette pratique, qui consiste à ne multiplier que les clones d'intérêt agronomique, a été adoptée en France après la seconde guerre mondiale, pour répondre aux objectifs quantitatifs et sanitaires de la production viticole. L'IFV (Institut Français de la Vigne et du vin), qui regroupe l'ENTAV (Etablissement National Technique pour l'Amélioration de la Viticulture) et l'ITV (Institut Technique de la Vigne et du Vin), est

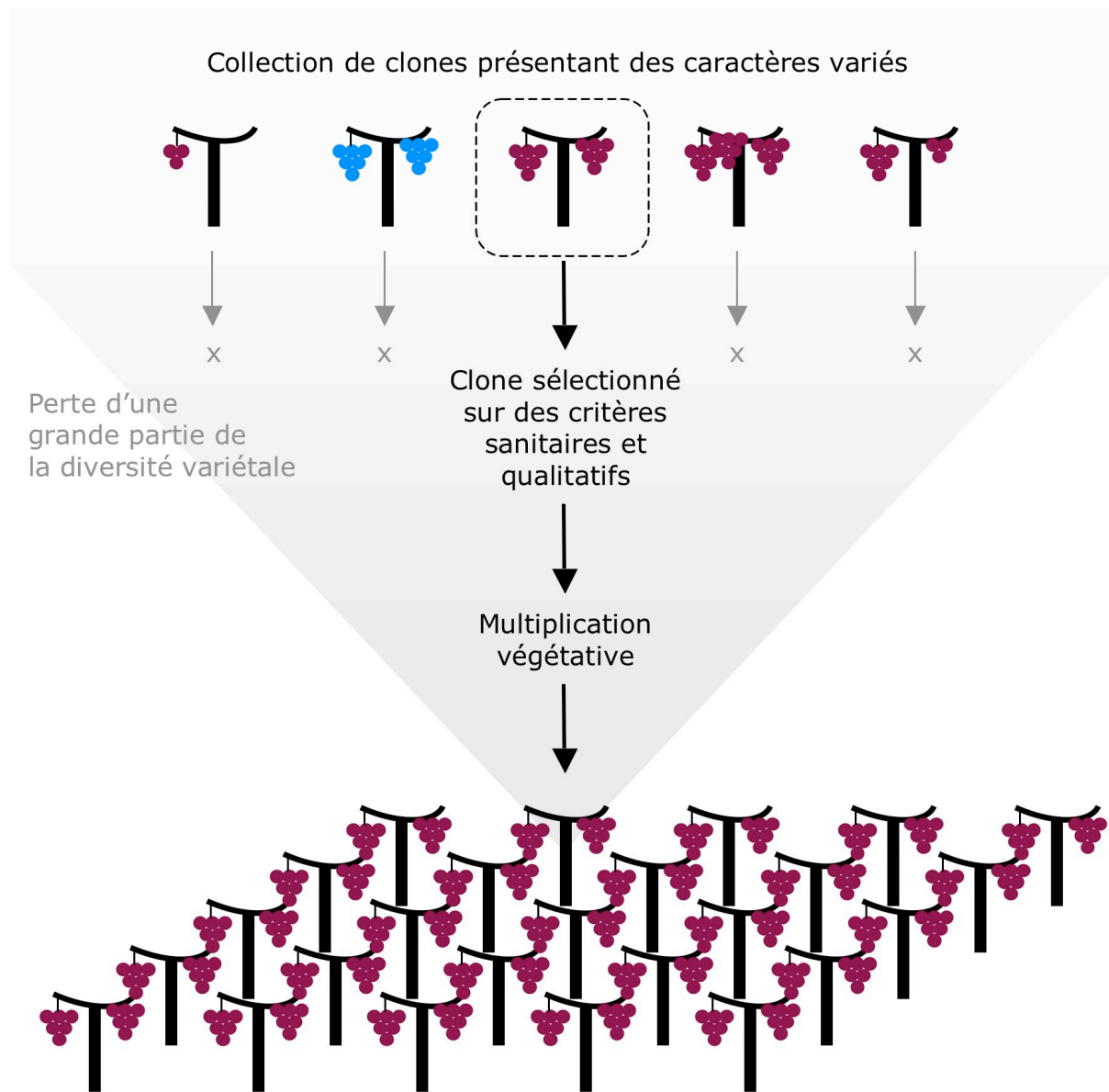


Figure 4 : Principe de la sélection clonale.

Nom commun	Nom scientifique	Taille du génome nucléaire (Mb)
Blé	<i>Triticum aestivum</i>	15 966
Pois	<i>Pisum sativum</i>	3 947
Maïs	<i>Zea mays</i>	2 292
Tomate	<i>Lycopersicon esculentum</i>	907
Vigne	<i>Vitis vinifera</i>	487
Peuplier	<i>Populus trichocarpa</i>	485
Riz	<i>Oryza sativa</i>	430

Tableau 1 : Taille de différents génomes de plantes.

désormais chargé de la coordination technique de la sélection de clones au niveau sanitaire (viroses et maladies) et génétique (rendement, qualité et typicité).

1.1.6 - Le génome de la vigne

Constatant les avancées agronomiques et scientifiques résultant du séquençage des trois premiers génomes de plantes, *A. thaliana*, le riz et le peuplier, le consortium International Grape Genome Program (IGGP) a été créé. Son but est de coordonner le développement des ressources génomiques sur la vigne et d'accélérer l'acquisition de connaissances sur ses caractères agronomiques, en permettant des approches de génomique fonctionnelle. Un projet de séquençage du génome de la vigne a ainsi été mis en place en 2004. A partir de juillet 2005, des équipes françaises et italiennes ont collaboré au séquençage du génome de la vigne par une stratégie de séquençage global (Whole Genome Shot-gun) visant à couvrir 12 équivalents du génome puis à les assembler.

Cependant, les variétés de vigne présentent un taux d'hétérozygotie d'environ 75% (Hocquigny *et al.*, 2004), ce qui aurait pu rendre l'assemblage et l'orientation des séquences difficiles. Pour contourner ce problème, le séquençage a été réalisé sur une accession appelée « PN40024 », qui présente un taux d'homozygotie proche de 93%. Cette accession a été créée à l'INRA de Colmar en réalisant plusieurs cycles d'autofécondation à partir du Pinot noir (Bronner and Oliveira, 1990).

Les premières données de séquençage du génome de la vigne ont été disponibles au milieu de l'année 2007. Leur analyse a montré que trois génomes « ancestraux » ont contribué à la formation du génome actuel de la vigne, qui présente une taille de 487 Mb (Tableau 1) et dont la séquence est composée à environ 41,4% de séquences répétées (Jaillon *et al.*, 2007).

Parallèlement à cette initiative, un autre laboratoire italien s'est lancé dans le séquençage du génome complet de la variété Pinot noir. Si les premières données ont été disponibles à la fin de l'année 2006, la taille des séquences assemblées (contigs) a rapidement cessé d'évoluer. Il semble en effet que le taux d'hétérozygotie du génome de cette variété ait considérablement ralenti l'assemblage des données.

I.2 - Constitution des cépages et variétés actuels

I.2.1 - Notions de cépage, de variété et de clone

a) Le cépage

La notion de cépage a émergé au Moyen Age et sa première citation écrite date de 1573. Le terme de cépage possédait alors un sens collectif plus proche du terme « encépagement », qui correspond à l'ensemble des cépages constituant les vignobles d'une région donnée. Au XVIIIe siècle, le cépage représentait l'unité culturelle de la vigne. Il s'agissait d'abord d'une unité taxonomique empirique qui se présentait comme « *une collection plus ou moins riche en clones plus ou moins voisins les uns des autres* » (Levadoux, 1956). En 1971, ce même auteur a proposé que les variétés cultivées reçoivent le nom de « cépage », qui était utilisé par les viticulteurs pour désigner un ensemble composé d'unités élémentaires, les clones. Le cépage a ensuite été considéré comme un ensemble de clones très proches les uns des autres, au point d'avoir été confondus sous un même nom (Galet, 1993). Cette définition se rapproche donc plus du terme botanique « cultivar » qui correspond à un individu issu d'un pépin, multiplié ensuite végétativement. Une des dernières définitions décrit le cépage comme « *une unité taxonomique propre à *Vitis vinifera* L. et qui est le produit d'un semis ou d'un individu unique au départ, multiplié par voie végétative* » (Boursiquot and This, 1999).

b) La variété

D'après le Comité Technique Permanent de la Sélection (CTPS), la variété se définit comme « *un ensemble homogène de matériel végétal sur la base des critères retenus pour la DHS (Distinction Homogénéité Stabilité), qui possède des caractéristiques ampélographiques homogènes avec des différences mineures au sein de la variété, permettant de différencier chaque individu d'une variété de ceux d'une autre variété, certaines caractéristiques agronomiques et technologiques stables et homogènes sur l'ensemble de la variété, des marqueurs moléculaires stables au sein d'une variété, avec cependant une possibilité de variation due à des mutations somatiques* ». En plus des critères morphologiques, les critères de classification en variétés reposent également sur l'intérêt technologique de la nouvelle

variation (couleur de la baie, de la pulpe, particularité aromatique...). Selon les dernières définitions, la variété se situerait donc au même niveau que le cépage.

c) Le clone

Le clone est une « *descendance végétative d'une variété conforme à une souche de vigne choisie pour son identité variétale, ses caractères phénotypiques et son état sanitaire* » (définition CTPS). Les clones de vigne portent le même nom de variété ou de cépage, et partagent normalement des caractères ampélographiques communs.

1.2.2 - Modes de propagation de la vigne

a) La reproduction sexuée

La reproduction sexuée résulte de la fusion des gamètes issus de deux individus dans le cas de l'allofécondation, ou du même individu lorsqu'il s'agit d'une autofécondation. Dans une population issue de la reproduction sexuée, la variabilité génétique a deux origines : la mutation et recombinaison génétique. La mutation provoque l'apparition de la majorité des formes alléliques présentes dans la population. Chez les Eucaryotes, la recombinaison implique deux mécanismes. Le premier est le crossing-over, c'est-à-dire l'échange de matériel génétique entre deux chromosomes homologues. Le second mécanisme, décrit dans la 2^e Loi de Mendel, est la distribution aléatoire des chromosomes homologues de chaque paire dans les gamètes. La recombinaison génétique qui a lieu lors de la reproduction sexuée peut générer un nombre considérable de nouvelles combinaisons alléliques et participe, par cet apport de diversité génétique, à l'adaptation et à l'évolution des espèces.

Chez la vigne cultivée, les pépins issus de la reproduction sexuée représentent la première forme de conservation. Cependant, ce procédé donne une descendance non identique au matériel de départ, un résultat que les professionnels du milieu viticole ne souhaitent que très rarement, puisqu'ils cherchent au contraire à multiplier à l'identique une plante qui présente des qualités d'intérêt agronomique. La variabilité générée au cours de la reproduction sexuée est principalement due à la forte hétérozygotie génétique de la vigne (environ 75%) qui favorise la distribution des caractères dans la descendance de la plante. Au cours de l'histoire de la vigne, la diversité générée par ce mode de multiplication a été valorisée par la recherche

de géotypes adaptés à de nouvelles conditions environnementales. Par exemple, le Cabernet sauvignon est issu du croisement entre le Cabernet franc et le Sauvignon blanc (Bowers and Meredith, 1997), alors que le croisement entre le Pinot et le Gouais a donné, en autres, les variétés Chardonnay, Gamay et Auxerrois (Bowers *et al.*, 1999). La reproduction sexuée n'est désormais utilisée que dans les programmes d'amélioration variétale.

b) La multiplication végétative

La simplicité technique de la multiplication végétative permet d'obtenir rapidement des plants conformes à la souche de départ, ce qui a probablement contribué au succès de la domestication de la vigne. Ce mode de propagation fut maîtrisé dès l'Antiquité (Billiard, 1913). Les agronomes grecs et romains décrivaient déjà de nombreuses techniques de bouturage, provignage et même de greffage. Différents procédés de conservation, comme le stockage en terre ou dans des jarres, étaient employés pour préserver les boutures de la dessiccation depuis leur prélèvement jusqu'à leur plantation. De nombreuses techniques sont toujours employées pour multiplier végétativement les espèces cultivées, par exemple le bouturage qui consiste à mettre en terre un fragment de plante dépourvu de racines. Des boutures de tiges (géranium), de rameaux (peuplier), de feuilles (saintpaulia) ou encore de racines (pissenlit) peuvent être utilisées. Le marcottage est un cas particulier de bouturage puisqu'un rameau de la plante est couché sous terre et reste lié à la plante mère jusqu'à la formation de racines. Certaines espèces, comme le fraisier (stolon) et le framboisier (drageon), utilisent la propagation végétative à l'état naturel. C'est également l'unique moyen de propagation des variétés de vigne apyrènes qui ne produisent pas de graines.

Plusieurs techniques utilisées actuellement ne varient guère de celles utilisées jusqu'à la fin du XIXe siècle, date à laquelle l'invasion du phylloxéra, provoqué par un puceron qui s'attaque principalement aux racines, a nécessité le greffage des variétés européennes sur des porte-greffes d'espèces américaines résistantes à ce parasite. Le greffage consiste à implanter un greffon, constitué d'un bourgeon ou d'un fragment d'organe portant des bourgeons, sur un porte-greffe (Figure 5). Le phylloxéra étant toujours présent dans le vignoble européen, le greffage sur des porte-greffes résistants est devenu le seul procédé permettant la culture de la vigne.

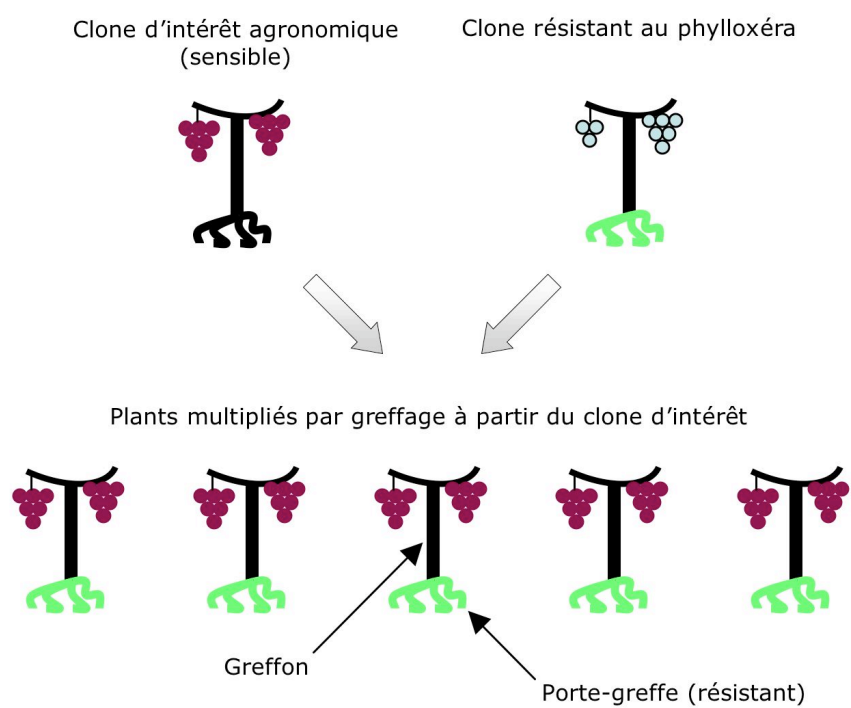
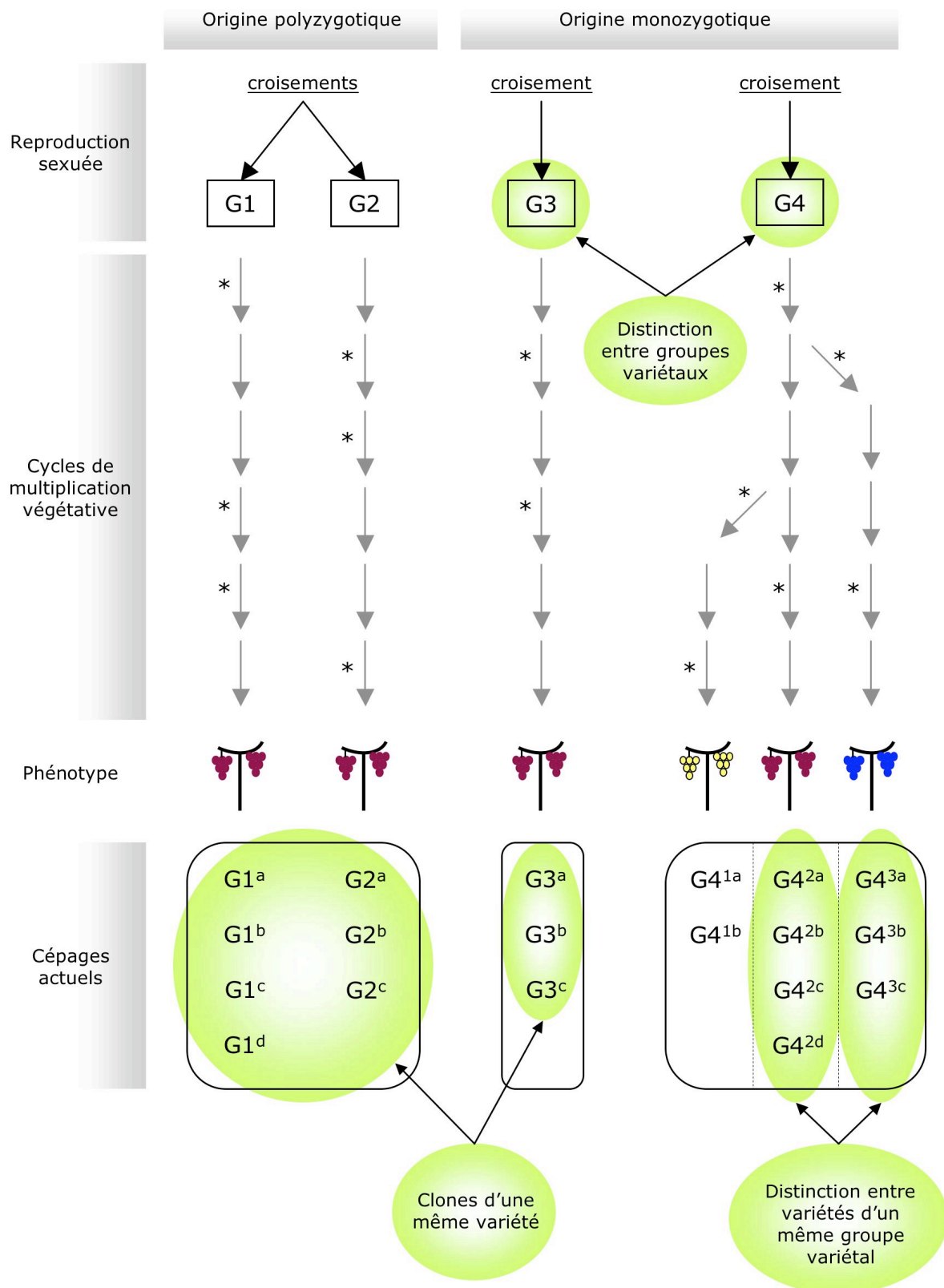


Figure 5 : Technique du greffage utilisée dans le cadre de la lutte contre le phylloxera.



G = Génotypes différents issus d'un ou de plusieurs événements de reproduction sexuée. ^a, ^{1a}, ... = divergence entre lignées somatiques par l'accumulation de mutations. * = Mutations différentielles accumulées lors des cycles de réplication végétative

Figure 6 : Représentation des différents scénarios possibles lors de la constitution des cépages.

c) Conséquences génétiques de la multiplication végétative

Par opposition à la multiplication par voie sexuée, qui entraîne la ségrégation des caractères, la multiplication végétative garantit en théorie l'identité au matériel de départ. Les descendants, possédant théoriquement le même patrimoine génétique, présentent généralement les mêmes caractéristiques agronomiques que la plante dont ils sont issus. L'apparition de mutations somatiques représente l'unique source de variation génétique chez les plantes à multiplication végétative. Pour les espèces fortement hétérozygotes comme la vigne, la multiplication végétative est donc un moyen efficace pour l'homme de fixer rapidement les caractères jugés intéressants.

I.2.3 - Scénarios pour la constitution des cépages

Plusieurs scénarios sont envisageables pour expliquer la constitution d'un cépage. Ils sont basés sur l'origine poly- ou monozygotique des individus qui le composent (Figure 6).

a) Origine polyzygotique

Le cépage est composé d'un mélange d'individus issus, par multiplication végétative, de plusieurs zygotes de départ descendants ou non du même croisement. Les clones ont été regroupés sous un même nom de cépage car ils possèdent des caractéristiques phénotypiques très proches (génotype G1-G2 dans la Figure 6). La variabilité des clones de ce cépage est due aux effets de la reproduction sexuée et à l'accumulation de mutations somatiques au cours des cycles de multiplication végétative.

b) Origine monozygotique

Les clones d'un cépage sont tous issus d'un zygote ancestral unique qui a été multiplié par voie végétative. Cette origine a été démontrée pour plusieurs variétés de vigne. Ainsi, les clones sont morphologiquement très proches mais peuvent présenter de petites différences génétiques liées aux mutations accumulées au cours des cycles de propagation (génotype G3 dans la Figure 6). Ces clones forment un ensemble phénotypique homogène et sont donc regroupés sous un même nom de cépage. Ce scénario peut s'appliquer aux créations variétales récentes, lors desquelles les cycles de multiplication végétative n'ont pas fait apparaître de formes phénotypiques pouvant avoir des usages technologiques différents.

Parfois, les mutations apparues au cours de la phase de propagation végétative provoquent des différences morphologiques marquées (génotype G4 dans la Figure 6). Lorsque la variation affecte un caractère technologique ou morphologique important, par exemple la variation de la couleur de la baie ou la teneur en composés aromatiques, les clones peuvent être distingués par des noms de cépages différents.

Les clones d'une même variété peuvent donc présenter des caractères différents, comme l'illustre la différence de taille de grappes observées entre deux clones de Cabernet sauvignon cultivés au Chili (Figure 7, Patricio Hinrichsen, INIA, Chili), ou la variabilité des teneurs en composés aromatiques détectée entre clones de Chardonnay (Figure 8, Eric Duchène, INRA, communication personnelle). La variété Pinot en est également une bonne illustration. Bien que son origine monozygotique ait été démontrée (Hocquigny *et al.*, 2004), cette variété est composée des Pinot noir, Pinot blanc, Pinot gris, Pinot meunier et Pinot moure qui présentent des différences phénotypiques flagrantes (Figure 9). Parmi 477 clones regroupant ces quatre cépages, 211 clones n'ont pas pu être différenciés à l'aide de 49 marqueurs microsatellites, révélant une grande homogénéité génétique au sein de la variété (Hocquigny *et al.*, 2004). De la diversité génétique est également présente chez les Cabernet sauvignon. Issus d'un croisement initial réalisé au XVIIe siècle, les descendants ont été plantés en France puis dispersés dans le monde, notamment au Chili. L'accumulation de mutations permet aujourd'hui de différencier les clones français et chiliens séparés il y a 150 ans (Figure 10) (Moncada *et al.*, 2006).

Si la diversité phénotypique observée entre les clones constituant un même cépage est essentiellement due à la variabilité génétique, son maintien est souvent le résultat de l'intervention de l'homme. L'amélioration variétale, c'est-à-dire la domestication de génotypes issus de populations sauvages locales ou la sélection des produits de croisements entre vignes cultivées et/ou vignes autochtones, mais aussi la sélection de clones affectés par des mutations somatiques sont à l'origine de la distinction de nouveaux cépages. Les plantes présentant les caractéristiques les plus intéressantes pour la diversification de la production et l'adaptation de la vigne aux terroirs ont été sélectionnées de façon empirique, puis propagées et diffusées par multiplication végétative, pérennisant ainsi les combinaisons alléliques et permettant la stabilisation de leurs caractères.

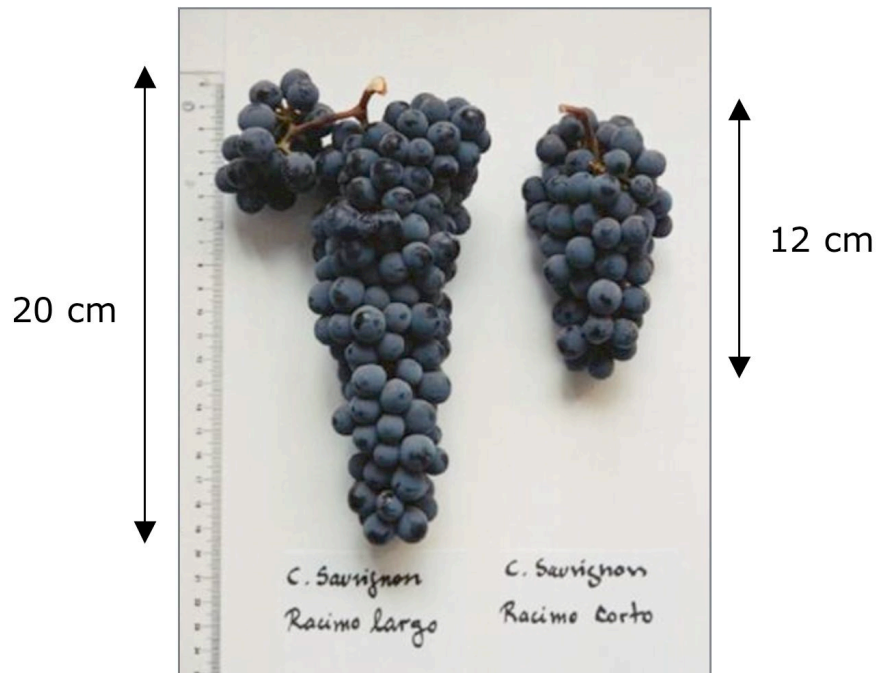


Figure 7 : Différence entre les tailles des grappes chez deux clones de cabernet sauvignon chiliens (P. Hinrichsen, INIA, Chili).

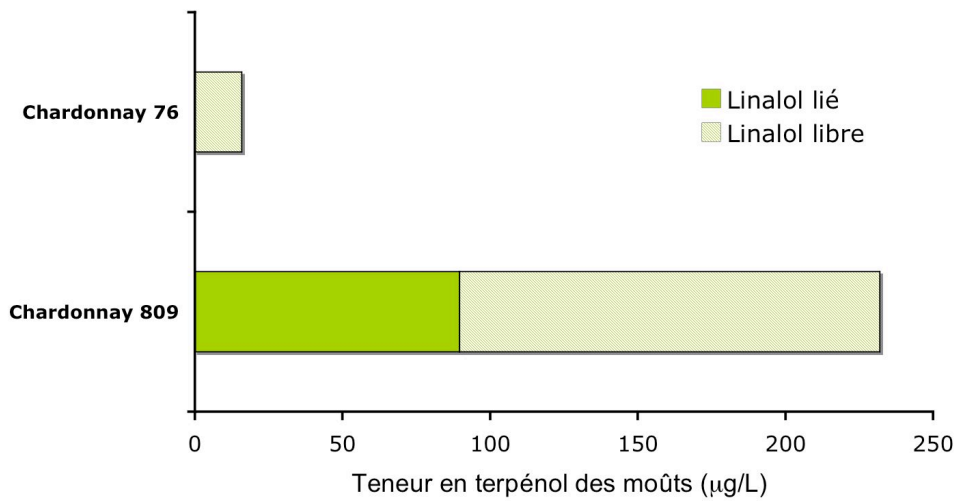
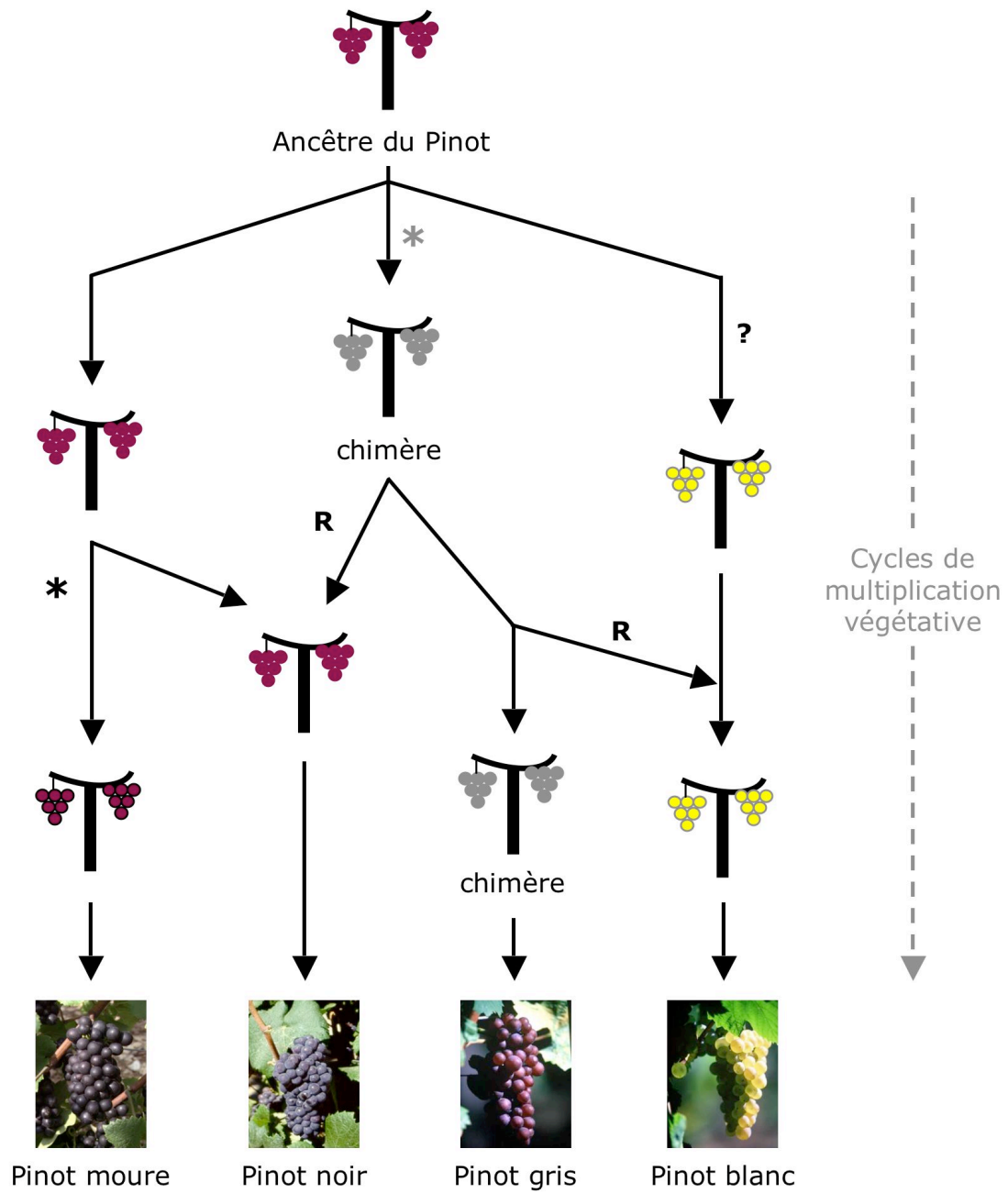


Figure 8 : Comparaison de la teneur en composés aromatiques (terpènes) mesurée sur deux clones de la variété Chardonnay (Eric Duchène, INRA, communication personnelle)



* = Mutation somatique, R = Réarrangements de tissus cellulaires

Figure 9 : Scénarios possibles de la diversification des Pinots (d'après Hocquigny *et al.*, 2004).

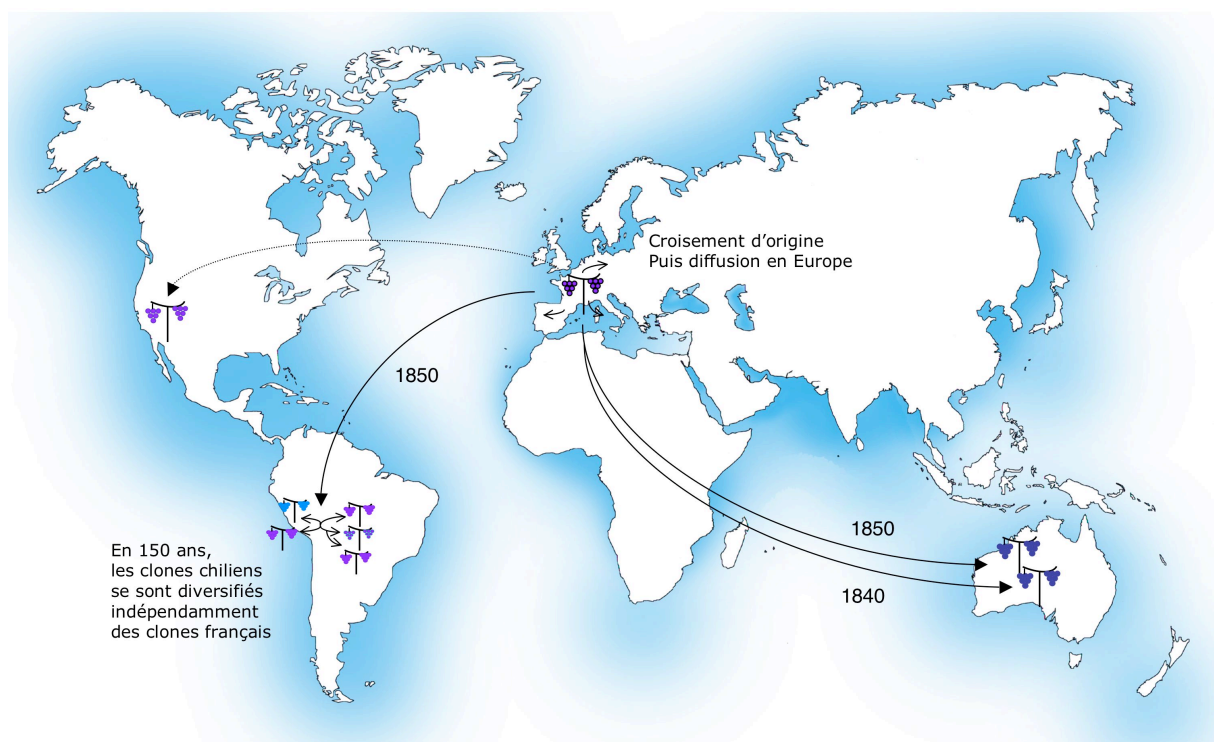


Figure 10 : Dispersion et diversification des clones de Cabernet sauvignon (d'après Moncada *et al.*, 2006).

Espèce	Nom commun	Forme de vie	Taux de mutation
10 espèces		Annuelles	0,0001 - 0,0004
<i>Mimulus gutatus</i>	(minule tachetée)	Annuelle	0,00045
<i>Kandelia candel</i>		Pérenne à cycle de vie court	0,0011 - 0,0012
<i>Rhizophora mangle</i>	(palétuvier rouge)	Pérenne	0,0021 - 0,0055
<i>Pinus sylvestris</i>	(pin sylvestre)	Pérenne	0,0042

Tableau 2 : Taux de mutation par génome haploïde et par génération, pour la déficience en chlorophylle chez différentes espèces végétales annuelles et pérennes (d'après Klekowski, 1998).

I.3 - Origines des mutations somatiques accumulées au cours des cycles de multiplication végétative

I.3.1 - Accumulation de mutations somatiques

Bien que la multiplication végétative assure en théorie le maintien de l'identité génétique de la plante mère à la plante fille, des mutations somatiques peuvent apparaître au cours des cycles de multiplication végétative, s'accumuler et ainsi générer une variabilité de type mutationnelle. L'accumulation de mutations somatiques au cours du développement de la plante produit une hétérogénéité génétique au sein même de l'individu.

Selon certains auteurs, les plantes pourraient avoir un taux de mutation par génération (sexuée ou asexuée) qui augmente avec la durée de vie de la plante. Différentes espèces végétales annuelles et pérennes ont été utilisées pour mettre en évidence la relation entre le taux de mutation et la durée de vie de l'individu. Les taux de mutation pour la déficience en chlorophylle ont été calculés pour différentes espèces annuelles appartenant à 10 genres de 5 familles de monocotylédones et de dicotylédones (Klekowski, 1998). Malgré une différence importante dans les tailles de génomes étudiés (de 0,2 pg/C à 5,5 pg/C), le taux de mutation évalué est relativement constant, puisqu'il est compris entre 1 et 4×10^{-4} par génome et par génération. Cependant, les espèces pérennes ont montré des taux de mutation 10 à 20 fois supérieurs à ceux des espèces annuelles (Tableau 2). Il semble donc exister une corrélation positive entre le cycle de vie (le nombre de divisions mitotiques entre la formation du zygote et la méiose) et le taux de mutation par génération.

De nombreuses observations de mutations somatiques ont été décrites chez les plantes cultivées et ornementales propagées de manière végétative. En 1907, Cramer a estimé que sur les 8800 variétés cultivées en Europe en 1899, 5000 d'entre elles provenaient de mutations somatiques (Shamel and Pomeroy, 1936). Beaucoup de ces mutants sont devenus des plantes économiquement importantes, comme le pamplemousse rose qui a été découvert en 1906 sur une seule branche mutée (Hartmann and Kester, 1975). Les mutations ont une influence considérable sur l'évolution des plantes supérieures.

1.3.2 - Origines des mutations somatiques chez les plantes à multiplication végétative

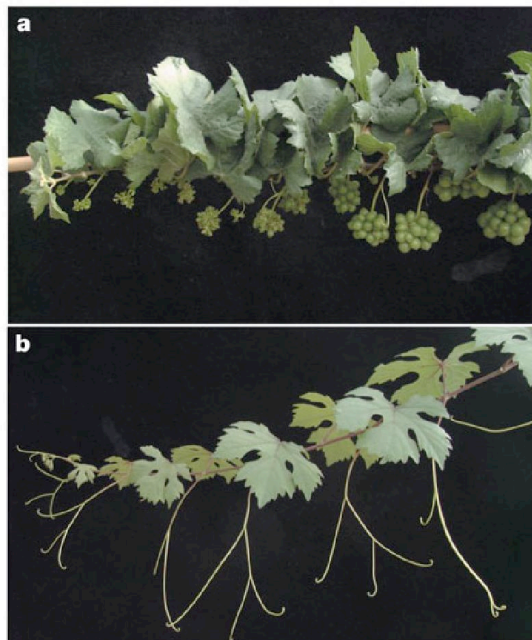
Au cours de la croissance végétative, plusieurs mécanismes peuvent générer de la variabilité génétique au sein d'un individu : des mutations ponctuelles durant la réplication de l'ADN, des insertions ou délétions, des changements dans le nombre ou dans la structure des chromosomes, des réarrangements cellulaires, des recombinaisons mitotiques, de la conversion génique, des duplications de gènes ou encore l'activité d'éléments transposables.

a) Mutations ponctuelles, insertions et délétions

Les mutations sont dites ponctuelles lorsqu'il s'agit de la substitution d'une base par une autre. Chez les Poacées, le taux moyen de substitution sur la région *Adh1-Adh2* a été estimé à $6,5 \times 10^{-9}$ substitutions par site et par an (Gaut *et al.*, 1996). Le génome effectif, qui correspond aux régions codant pour des protéines, ne représente pas la majorité du génome. Ainsi, la plupart des mutations sont strictement neutres car elles affectent des régions non codantes. Le taux de mutation par génome effectif et par génération sexuée a été estimé à 0,14 chez la drosophile et à 1,6 chez l'homme (Drake *et al.*, 1998). Du fait de la dégénérescence du code génétique, environ un tiers des mutations ponctuelles sont silencieuses. Les mutations qui ont un effet sur le phénotype sont majoritairement des mutations faiblement délétères et, pour être transmises, elles ne doivent pas affecter trop fortement la fitness de l'organisme ou de la cellule mutée, c'est-à-dire sa capacité à survivre et à se reproduire.

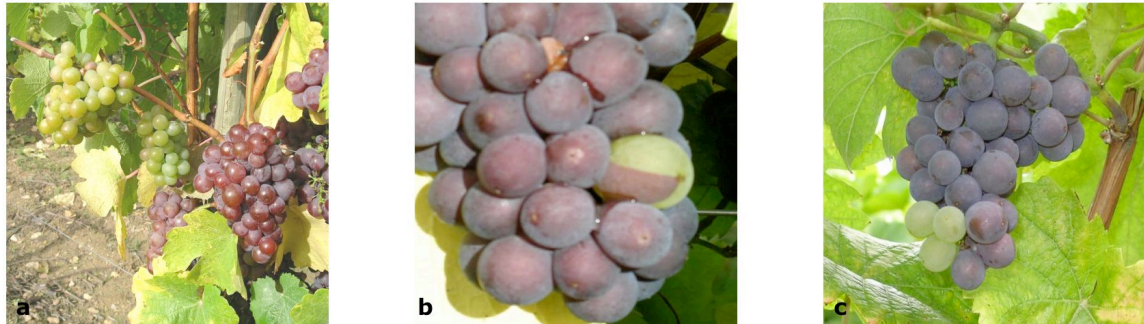
Un exemple de mutation ponctuelle affectant le phénotype est décrit chez le Pinot meunier (Boss and Thomas, 2002). Cette variété de Pinot s'est différenciée à partir du Pinot noir (Figure 9). Une mutation ponctuelle de l'allèle *VvGAI1*, apparue dans la couche cellulaire L1 du Pinot meunier, provoque le remplacement d'un acide aminé Leucine (petit et hydrophobe) par un acide aminé Histidine (grand et basique) dans la protéine correspondante (Figure 11). Ce changement altère les propriétés de réponse de la protéine à la gibbérelline, et provoque le phénotype « nain » chez les plantes régénérées par embryogenèse somatique à partir de cette couche cellulaire. Les plantes régénérées à partir de la couche L2 présentent les mêmes caractéristiques que le Pinot noir (Figure 11).

VvGAI1 SPTGKGKMWD ADPQQDAGMD EL **L**AVLGYNV KASDMAEVAQ KLEQLEEVIV
Vvgai1 SPTGKGKMWD ADPQQDAGMD EL **H**AVLGYNV KASDMAEVAQ KLEQLEEVIV



Alignement des séquences protéiques déduites des allèles *VvGAI1*. Le changement du résidu L en H provoque le phénotype « nain » chez les plantes régénérées à partir de la couche cellulaire L1 (a), alors que les plantes L2 sont normales (b).

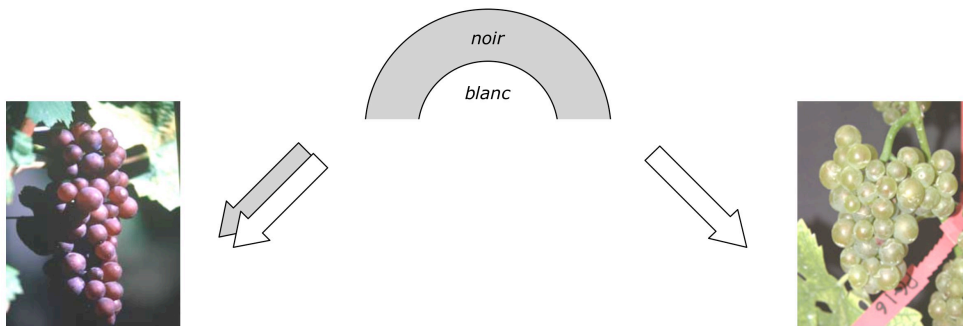
Figure 11 : Exemple d'une mutation ponctuelle affectant le phénotype chez la vigne : le cas du Pinot meunier (d'après Boss et Thomas, 2002).



a) Grappes de raisins blancs et gris présentes sur le même pied de vigne de Pinot gris. b et c) Baie et grappe sectorisées.

Figure 12 Exemples de structures chimériques observées chez la vigne.

Le Pinot gris est une chimère composée de deux couches tissulaires génétiquement différentes : l'épiderme est « noir » alors que les couches internes sont de type « blanc »



La propagation végétative maintient la structure à l'état chimérique

La reproduction sexuée (ici par autofécondation) ne transmet que le caractère « blanc »

Figure 13 : Une conséquence de la multiplication végétative : la propagation de l'état chimérique du Pinot gris (d'après Hocquigny *et al.*, 2004).

Des insertions ou délétions d'une ou de plusieurs bases, ou encore des inversions peuvent également se produire. Chez la vigne, un grand nombre de différences clonales ont été mises en évidence à l'aide de marqueurs microsatellites (Sefc *et al.*, 2001). Ces marqueurs moléculaires sont très polymorphes chez la vigne hétérozygote et se sont révélés efficaces pour analyser la structure génétique et la différenciation des vignes cultivées (Aradhya *et al.*, 2003), pour identifier des variétés (This *et al.*, 2004), pour établir des relations phylogéniques (Bowers *et al.*, 1999) et pour construire des cartes génétiques (Adam-Blondon *et al.*, 2004; Doligez *et al.*, 2006).

b) Les changements chromosomiques

Les changements chromosomiques structuraux peuvent être des délétions, des duplications, des inversions ou des translocations de parties de chromosomes. Les délétions ont souvent des effets négatifs, comme la perte de certains gènes ou l'instabilité des chromosomes sur l'extrémité tronquée. Les inversions peuvent perturber le taux ainsi que l'ordre de transcription des gènes. Les translocations intra- ou interchromosomiques, comme d'autres types de mutations, peuvent perturber l'appariement des chromosomes. Le changement du nombre de chromosomes peut, quant à lui, impliquer des jeux complets de chromosomes (polyploïdie) ou partiels (aneuploïdie). La polyploïdie est un phénomène courant chez les plantes et a certainement joué un rôle important dans la spéciation et l'évolution d'espèces cultivées. Les plantes polyploïdes sont souvent plus grandes et plus vigoureuses, ce qui leur confère un avantage pour l'invasion de nouvelles niches écologiques ou la sélection par l'homme. Par exemple, les vignes auto-tétraploïdes, présentant des baies plus grosses et des grappes plus lâches, ont attiré l'attention des sélectionneurs de raisins de table (Alleweldt and Possingham, 1988).

c) Chimères et réarrangements cellulaires

Lorsqu'une mutation affecte la cellule d'un méristème apical, cette cellule peut se diviser par mitose et produire un secteur muté, génétiquement différent du reste de la plante. Des réarrangements cellulaires, entre les couches cellulaires (L1 et L2) d'un méristème, peuvent conduire à la formation de plantes ou parties de plantes chimériques (Figure 12). La reproduction sexuée, qui produit des descendants via une cellule unique par mitose, méiose ou syngamie, restaure une homogénéité génétique dans la descendance. A l'inverse, la multiplication végétative produit des descendants via des éléments de propagation constitués

de plusieurs cellules (greffons, boutures...) et une organisation chimérique peut donc être propagée lors de bouturages, de greffages ou de divisions, si l'élément de propagation contient des tissus génétiquement différents.

La détection simultanée de trois allèles microsatellites au même locus, dans les clones polymorphes d'une collection de 145 accessions appartenant au Pinot gris, Pinot noir, Pinot blanc, Pinot meunier et Pinot moure, a démontré la prédominance du chimérisme chez les clones de vigne (Hocquigny *et al.*, 2004). Le Pinot gris est, par exemple, composé d'un épiderme de phénotype « noir » (probablement présent chez le pinot ancestral, voir Figure 9) alors que la couche cellulaire interne est génétiquement de type « blanc ». Les plantes régénérées par embryogenèse somatique à partir des couches cellulaires L1 ou L2 présentent des phénotypes différents : l'état chimérique est transmis lors de la réplication végétative, alors que la reproduction sexuée transmet uniquement le caractère « blanc » à la descendance (Figure 13) (Hocquigny *et al.*, 2004).

d) L'activité des éléments transposables

Les éléments transposables sont des éléments génétiques mobiles, capables de se multiplier dans le génome, parfois de manière autonome. Ils ont été trouvés dans tous les génomes dont ils sont parfois les constituants majeurs (Feschotte *et al.*, 2002). Le nombre et le caractère invasif des éléments transposables en font des acteurs majeurs de la structure, de l'expression et de l'évolution des génomes.

Jusqu'à présent, seuls trois éléments transposables, des rétrotransposons à LTR (Long Terminal Repeats), ont été décrits chez la vigne. L'élément *Vine-1* a été découvert dans le gène *Adhr* (Verries *et al.*, 2000), la famille *Tvv1* a été reconstituée grâce à une stratégie de marche sur chromosome (Pelsy and Merdinoglu, 2002), et l'élément *Gret1* a été détecté en amont du gène *VvmybA1*, où son insertion, en empêchant l'expression du gène, aurait conduit à l'apparition des baies de couleur blanche chez les pinots (voir page 55) (Kobayashi *et al.*, 2004).

Les éléments transposables, plus particulièrement la classe des rétrotransposons à LTR, leur mécanisme de transposition, ainsi que leurs implications dans la structure et la dynamique des génomes sont présentés dans le chapitre suivant.

Chapitre II : Les éléments transposables

II.1 - Evolution des pensées

Les débuts de la génétique

En 1858, Darwin formule la théorie de l'évolution selon laquelle tous les êtres vivants se développent à partir de formes de vie originelles. La loi du plus fort fait que les êtres qui s'adaptent le mieux aux conditions changeantes de l'environnement survivent et se multiplient, alors que les plus faibles disparaissent. Le moine J.G. Mendel, considéré aujourd'hui comme le père fondateur de la génétique, conclut en 1865, grâce à ses expériences de croisements entre petits pois, que les caractéristiques des plantes parentales sont transmises à leur descendance selon des lois pré-établies. La pertinence de ses travaux ne sera véritablement reconnue que plusieurs années plus tard. En 1909, le biologiste Johannsen donne aux « éléments » le nom de « gènes » (Johannsen, 1909). En parallèle, Morgan constate que les chromosomes des drosophiles sont porteurs des gènes et que beaucoup d'entre eux sont alignés sur les chromosomes dans un ordre bien précis (Morgan, 1910).

L'ADN, vecteur de l'information génétique héréditaire

Malgré l'accumulation croissante de preuves, notamment suite aux travaux de Griffith qui seront repris par Avery, la communauté scientifique n'accepte pas encore l'idée selon laquelle l'ADN est le support de l'hérédité (Griffith, 1928; Avery *et al.*, 1944). Selon les thèses les plus largement acceptées, l'ADN n'est qu'une molécule simple, et donc incapable de véhiculer une information complexe. En partie à cause de l'influence de la théorie tétranucléotidique proposée par Levene, l'un des plus grands spécialistes des acides nucléiques de l'époque, qui stipulait que la structure de l'ADN est régulière et monotone, les biologistes ont cru pendant des années que la structure des acides nucléiques était trop simple pour porter une information génétique très complexe. Les protéines, dont l'immense diversité avait été pressentie, semblaient à l'époque de bien meilleurs candidats pour véhiculer l'information génétique.

L'ADN a ensuite été décrit comme porteur d'une certaine spécificité, comme une molécule qui n'a pas une structure polymérique monotone et qui est donc susceptible de contenir une information. Les travaux de Hershey en 1946, puis ceux de Chargaff en 1950, contribuent

finalement à répandre l'idée que l'ADN peut être une molécule porteuse de l'information génétique et qu'il occupe une place centrale dans les mécanismes héréditaires (Hershey, 1946; Chargaff, 1950).

La structure de l'ADN, une étape décisive vers la compréhension des mécanismes de réplication

En 1953, deux jeunes chercheurs synthétisent les résultats de plusieurs expériences et publient un petit article d'une seule page qui allait marquer le monde de la génétique et de la biologie moléculaire. « *Nous souhaitons suggérer une structure pour l'acide désoxyribonucléique (ADN). Cette structure présente de nouvelles caractéristiques qui sont d'un intérêt biologique considérable* » (Watson and Crick, 1953). La détermination de cette structure en double-hélice par Watson et Crick a permis par la suite de comprendre l'ensemble des mécanismes moléculaires de l'expression génétique, comme la réplication de l'ADN, la transcription des ARN, le code génétique... et la formulation du « dogme central » selon lequel un gène est traduit sous forme de protéine en passant par une copie de ce gène (Gamow, 1954; Crick *et al.*, 1957).

La découverte des « éléments de contrôle » par Barbara McClintock

Au début des années 1940, la généticienne Barbara McClintock met en évidence chez le maïs une paire de loci qui semble impliquée dans l'apparition de mutations spontanées et réversibles à des allèles pourtant considérés comme stables (McClintock, 1941; McClintock, 1942; McClintock, 1944). En observant leurs effets sur les feuilles et les graines, elle pense dans un premier temps avoir perturbé les mécanismes de régulation de l'activité de gènes. Constatant les difficultés qu'elle rencontre pour les localiser sur les chromosomes, McClintock réalise quelques années plus tard qu'il ne s'agit pas de loci génétiques, mais d'éléments chromosomiques qui se déplacent dans le génome. Il devient alors évident pour elle que l'existence de ces nouveaux éléments n'est pas sans conséquences sur les gènes. Elle développe en 1950 une théorie qui explique comment les « *éléments de contrôle* » peuvent orchestrer le développement et la différenciation de l'organisme (McClintock, 1956a; McClintock, 1956b). Selon elle, la clef réside dans la coordination de la transposition d'un groupe d'éléments de contrôle qui transposent de concert, inhibant et modulant les effets des gènes pour exécuter le programme de développement de la cellule.

Les éléments mobiles confrontés à la définition de l'évolution

La découverte par Barbara McClintock des éléments transposables chez le maïs a révélé un aspect alors encore inconnu de la génétique. Cependant, elle n'a pas été immédiatement prise en considération et il faut attendre une dizaine d'années pour que ses travaux trouvent réellement un écho auprès de la communauté scientifique. En effet, pendant de nombreuses années, l'idée même de l'évolution repose sur un postulat selon lequel le génome et son organisation, structurale et fonctionnelle, sont le résultat de sélections positives. Selon cette idée, seuls les éléments positifs pour l'individu sont conservés dans le génome, puis répandus dans la population du fait de leur rôle bénéfique. Dans cette théorie, le rôle des éléments transposables semble donc difficile à concevoir.

Des « éléments régulateurs »

La compréhension du mécanisme de régulation des gènes est née du modèle « *opéron* », proposé en 1960 par Jacob et Monod pour expliquer le contrôle d'un ensemble de gènes structuraux par un groupe de gènes régulateurs adjacents (Jacob and Monod, 1965). Prenant connaissance de ce modèle, McClintock établit le parallèle entre les systèmes de contrôle observés chez le maïs et chez les bactéries, présentant pour la première fois ses éléments comme des « *opérateurs* » et des « *régulateurs* » (McClintock, 1961).

Les effets des éléments mobiles sur le génome et leur implication dans la régulation des gènes sont ensuite peu à peu pressentis. Dans les années 1960-1970, la caractérisation des motifs spécifiques qui permettent aux éléments de s'exciser et de se réinsérer dans le chromosome bactérien, la détection de ces mêmes éléments dans d'autres organismes, leur participation à la propagation rapide de la résistance aux antibiotiques, et plusieurs autres recherches mettent en évidence que le mouvement des éléments génétiques dans et entre les génomes est un mécanisme de très grande importance. En 1969, Britten et collaborateurs proposent que certains éléments répétés peuvent avoir un rôle en tant que « *regulatory network* » (réseaux de régulation) pour coordonner l'expression de gènes très différents et dispersés dans le génome (Britten and Davidson, 1969).

De l'égoïsme à l'inutilité...

Les théories de Dawkins et Kimura illustrent parfaitement la progression de la pensée scientifique vers l'idée que l'évolution et la sélection peuvent être envisagées à deux niveaux : celui de l'ADN et celui de l'individu. A l'échelle de l'ADN, la sélection positive des éléments

transposables peut résulter de leur capacité à se répliquer plus rapidement que les séquences de l'hôte. Du fait de leur implication dans certains cas de mutations délétères pour l'hôte, ils sont d'abord considérés comme des parasites génomiques, qui se maintiennent au cours de l'évolution grâce à leur énorme capacité de réplication. Cette théorie de l'« *ADN parasite* » est illustrée par l'hypothèse de Dawkins sur l'« *ADN égoïste* », selon laquelle la vie est simplement le moyen de propagation de l'ADN (Dawkins, 1976). Elle est ensuite discutée entre plusieurs auteurs qui considèrent l'ADN égoïste comme le « *parasite ultime* » et commencent à entrevoir son implication dans l'évolution du génome (Doolittle and Sapienza, 1980; Dover and Doolittle, 1980; Orgel and Crick, 1980; Orgel *et al.*, 1980). La découverte de McClintock trouve enfin un écho auprès de la communauté scientifique, comme le prouve ses nominations au prix Nobel entre 1976 et 1983, l'année où elle le remporte (Comfort, 2001). Cette même année, la théorie sur l'évolution neutraliste proposée par Kimura permet de reconsidérer les éléments transposables, puisqu'elle postule que ce ne sont pas les mutations les plus favorables qui se propagent dans la population, mais à l'inverse les mutations les plus néfastes qui sont contre-sélectionnées (Kimura, 1983). Cette théorie laisse donc entrevoir une place pour les éléments transposables puisque le génome doit pouvoir contenir un certain nombre de séquences inutiles ou ayant des effets peu délétères, sans qu'elles soient pour autant éliminées par la sélection. Un élément transposable peut certes être responsable d'effets délétères, mais malgré tout se maintenir ou se propager à l'échelle de la population. Les éléments transposables sont ainsi considérés comme de l'« *ADN poubelle* », simplement inutile, dans le meilleur des cas.

Une source d'adaptation pour l'hôte, un statut très récent

L'acquisition de nouvelles données a désormais fait passer les éléments transposables du statut peu enviable de « *junk DNA* » (qui peut être traduit par « ADN n'importe quoi ») à celui de « *seeds of evolution* » (une source d'évolution) (Brosius, 1991), certains auteurs les comparant même à une matière précieuse comme l'or (Volff, 2006). Sans aller jusque là, il est possible de considérer un continuum d'interactions entre ces éléments et leur hôte, allant du parasitisme à la symbiose (Kidwell and Lisch, 2001). En effet, les éléments transposables sont, par leur nombre et leur mobilité, un réservoir quasi-inépuisable de variabilité génétique, l'essence même de l'adaptabilité et de l'évolution des génomes. Une perspective que Barbara McClintock avait déjà évoquée lors de leur découverte dans les années 1940 (McClintock, 1984).

II.2 - Classification des éléments mobiles

Deux classes d'éléments transposables se distinguent par leur mode de transposition, les éléments de la classe I et ceux de la classe II, qui sont toutes les deux composées d'éléments autonomes et non-autonomes (Figure 14).

II.2.1 - Les éléments de classe I : les rétrotransposons

Les éléments de la classe I, ou rétrotransposons, ont un mécanisme de transposition de type « copier-coller » qui nécessite la transcription d'une molécule d'ARNm intermédiaire, puis sa transcription inverse en ADN. La copie maître du rétrotransposon reste à sa position initiale à l'issue du cycle de rétrotransposition, alors que les nouvelles copies s'insèrent à de nouveaux sites. Parmi les éléments de classe I, on peut distinguer les rétrotransposons et les rétrotransposons à LTR.

a) Les rétrotransposons à LTR

La région interne des rétrotransposons à LTR est bordée par deux répétitions orientées dans le même sens, les LTR (Long Terminal Repeats), qui contiennent les promoteurs, les terminateurs et les séquences régulatrices de la transcription (Figure 14). La séquence interne est généralement constituée de la séquence codante des gènes *gag* et *pol*, codant respectivement pour la nucléocapside (ou particule pseudo-virale, VLP) et pour le complexe enzymatique. Le complexe enzymatique *pol* comprend les domaines protéase (PR), transcriptase inverse (RT), RnaseH et intégrase (Int). Le cycle de rétrotransposition des rétrotransposons à LTR est présenté en détail dans le chapitre suivant (page 31).

La structure des rétrotransposons à LTR est très proche de celle des rétrovirus. Cependant, la séquence des rétrovirus contient un gène *env* qui code pour une enveloppe et permet aux particules virales d'infecter d'autres cellules. Ainsi, les rétrovirus ont la capacité de passer de cellule à cellule, ce qui les rend infectieux, alors que les rétrotransposons ont un cycle exclusivement intra-cellulaire. Dans certains cas, les rétrotransposons possèdent un cadre de lecture supplémentaire, semblable à celui du gène *env* des rétrovirus (Figure 14). C'est notamment le cas de l'élément *gypsy* du génome de la drosophile, qui présente des capacités

infectieuses (Kim *et al.*, 1994; Pelisson *et al.*, 2002). La question de la relation phylogénique entre les rétrotransposons et les rétrovirus a été posée par Howard Temin, qui a émis l'hypothèse selon laquelle les rétrovirus sont issus des rétrotransposons par l'acquisition du gène de l'enveloppe (Temin, 1980). A l'inverse, les rétrotransposons sont peut-être des rétrovirus dégénérés qui ont perdu leur gène de l'enveloppe et donc leur capacité à infecter de nouvelles cellules.

Ty1-copia-like et Ty3-gypsy-like

Deux types de rétrotransposons, *Ty1-copia* (ou *Pseudoviridae*) et *Ty3-gypsy* (ou *Métaviridae*), se distinguent par des caractéristiques structurales, notamment dans l'ordre des domaines fonctionnels de la région *pol* (Havecker *et al.*, 2004). Chez les éléments de type *Ty1-copia*, l'ordre des domaines fonctionnels est *gag*, protéase, intégrase, RT (transcriptase inverse) et RnaseH. Chez les éléments de type *Ty3-gypsy*, le domaine RT-RNaseH est situé entre les domaines protéase et intégrase (Figure 14). La séquence de ces rétrotransposons fait généralement plusieurs kilobases.

Les LARD

Les LARD (pour Large Retrotransposon Derivative) dérivent de rétrotransposons fonctionnels, mais leur région interne a accumulé des mutations et ne contient pas de cadre de lecture ouvert suffisamment grand leur permettant de synthétiser leurs propres protéines fonctionnelles. C'est le cas des éléments *Sukkula* (Manninen and Schulman, 1993) et *Dasheng* (Jiang *et al.*, 2002a; Jiang *et al.*, 2002b) détectés respectivement chez l'orge et le riz.

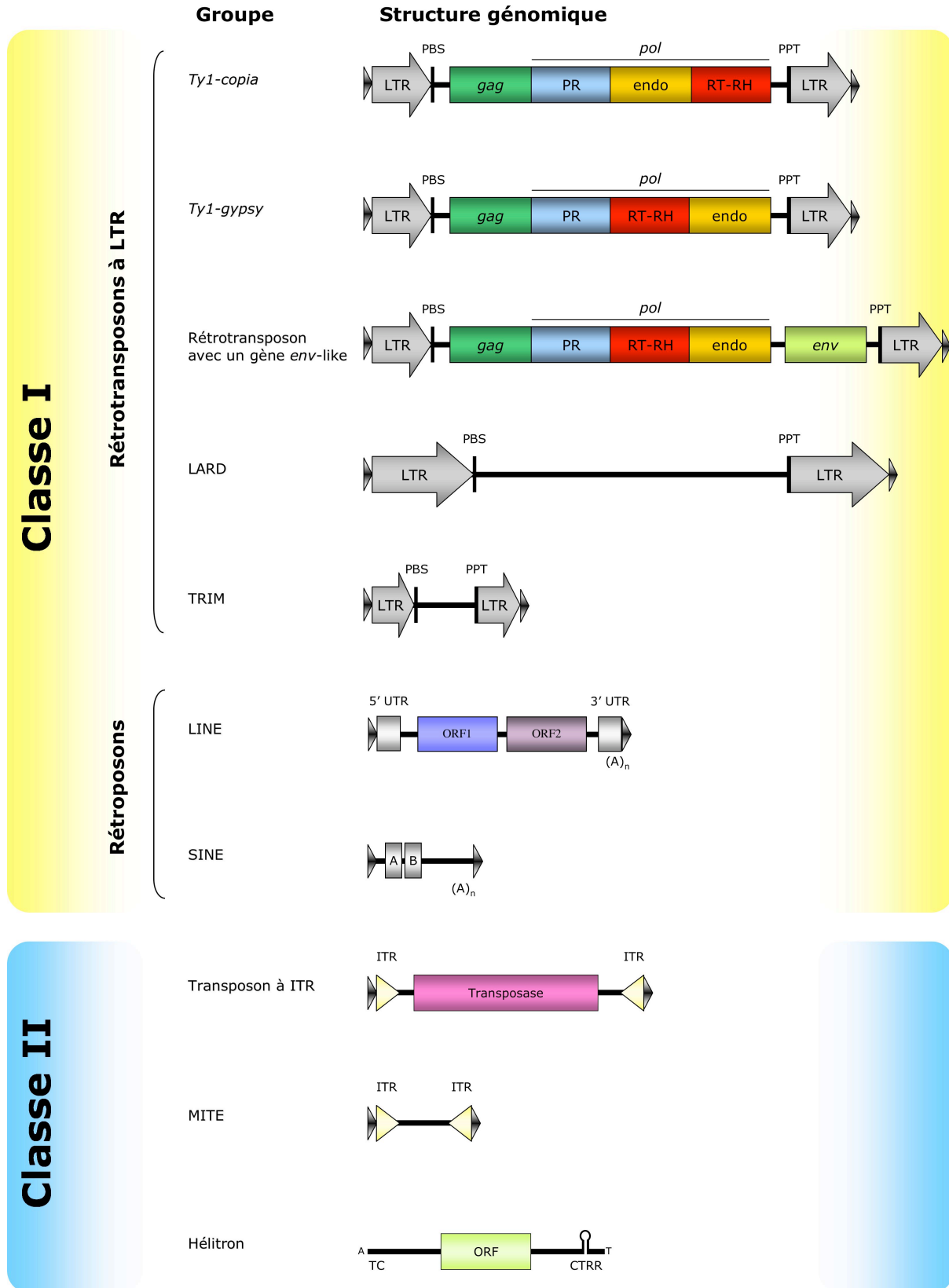
Les TRIM

Les TRIM (pour Terminal Repeat in Miniature) présentent les caractéristiques classiques des rétrotransposons à LTR, mais aucune région codante. Ils se différencient des éléments LARD par leur très petite taille (en général moins de 540 pb) et par des répétitions terminales (LTR) très courtes (environ 140 pb) (Witte *et al.*, 2001).

b) Les rétroposons (ou rétrotransposons sans LTR)

Les LINEs

Les LINEs (Long Interspersed Nuclear Elements) sont des éléments de grande taille (quelques kb) qui sont transcrits par l'ARN polymérase II. Ils codent pour la machinerie enzymatique



— = ADN non-codant. □ = ADN codant. ▶ = Duplication du site cible. LTR = Long Terminal Repeats. PBS = Primer Binding Site. PPT = Poly-Purine Tract. PR = Protéase. Endo = Endonucléase/intégrase. RT = Transcriptase Reverse. RH = RNaseH. Env = Gène de l'enveloppe. LARD = LARge Retrotransposon Derivative. TRIM = Terminal Repeat retrotransposon in Miniature. LINE = Long Interspersed Nuclear Element. SINE = Short Interspersed Nuclear Element. UTR = Untranslated Region. ORF = Open Reading Frame. (A)_n = Queue poly(A). ITR = Inverted Terminal Repeats.

Figure 14 : Classification générale des éléments transposables.

nécessaire à leur cycle de transposition. L'élément est d'abord transcrit à partir d'un promoteur interne présent dans la région non traduite UTR 3' (Untranslated Region), puis traduit pour donner, à partir des cadres de lecture ORF1 et ORF2 (Open Reading Frame), deux protéines nécessaires au cycle de rétrotransposition (Kazazian, 2004). L'activité intégrase de la protéine codée par l'ORF2 permet la coupure de l'ADN au niveau du site cible, à partir duquel la transcription d'un ARN messager est réalisée. L'ARN est ensuite reverse-transcrit en ADN complémentaire qui, à l'état simple-brin, est inséré au niveau du site cible. La synthèse du second brin est alors réalisée et donne une copie ADN double brin du rétroposon, bordée par deux duplications du site cible.

Les SINEs

Les SINEs (Short Interspersed Nuclear Elements) sont des éléments de petite taille (quelques dizaines ou centaines de bases) possédant un promoteur, mais pas de région interne (Figure 14). Le SINE le plus connu est l'élément *Alu* qui est présent en plus d'un million de copies dans le génome humain et représente plus de 10% de ce génome (The International Human Genome Sequencing Consortium, 2001). Ces éléments non-autonomes dépendent d'autres éléments pour leur transposition (voir page 29).

II.2.2 - Les éléments de classe II : les transposons

Les transposons à ITR

Les éléments de la classe II, les transposons, s'excisent de la séquence d'ADN et s'insèrent à un nouveau site, souvent proche de leur localisation initiale, par un mécanisme de type « couper-coller », c'est-à-dire sans intermédiaire ARN. La copie de l'élément est excisée de sa position originale dans le génome, puis insérée à un autre site cible. Le cycle des transposons est contrôlé par une transposase et est souvent couplé à une amplification de la copie initiale. La structure des transposons est composée de la séquence codante de la transposase, représentée par un ou plusieurs gènes, encadrée par des séquences répétées inversées (Inverted Terminal Repeats ou ITR) (Figure 14). Les transposons sont très largement distribués, notamment au sein des génomes eucaryotes, et leur taille dépasse rarement quelques kb. De nombreux transposons à ITR ont été décrits, comme les superfamilles *P* initialement détectées chez la drosophile (Daniels *et al.*, 1990), la famille *Tc1-mariner* (*Tc1* de *Caenorhabditis elegans* et *mariner* de la drosophile) (Plasterk *et al.*,

1999), les familles *Mu* et *CACTA* chez le maïs puis chez d'autres plantes (Gierl *et al.*, 1989; Chandler and Hardeman, 1992), et *hAT* (*hobo* de la drosophile, *Activator* du maïs et *Tam-3* d'*Anthirrinum*) (Kempken and Windhofer, 2001). Des éléments appartenant à ces familles ont depuis été décrits dans d'autres génomes de champignons, de plantes et d'animaux.

Les MITE

Des éléments de plus petite taille et très abondants ont été mis en évidence dans le génome du riz (Bureau *et al.*, 1996), puis dans ceux d'autres plantes, d'insectes et de mammifères. Ces éléments d'environ 600 pb, sont appelés MITE pour Miniature Inverted-repeat Transposable Elements (Figure 14).

Les Hélitrons

Les Hélitrons forment une classe à part chez les éléments transposables (Figure 14), puisque leur cycle de transposition ne fait pas appel à une transposase, mais à un mécanisme appelé « Rolling-circle Replication », de type cercle roulant (Kapitonov and Jurka, 2001). La réplication des Hélitrons utilise les protéines de réplication de l'hôte, des hélicases, pour la synthèse de la nouvelle copie.

II.2.3 - Vers une nouvelle classification ?

La classification des éléments transposables est régulièrement mise à jour, notamment après la découverte de nouveaux éléments. Il est en effet parfois difficile d'intégrer dans la classification en vigueur des éléments qui présentent des caractéristiques structurales nouvelles et des cycles de transposition jusqu'alors inconnus. En décembre 2007, une nouvelle classification a été proposée, qui distingue les éléments sur la base de leur mécanisme de transposition, notamment sur les étapes d'excision et d'insertion, sur la similarité de leurs séquences et sur leurs relations structurales (Wicker *et al.*, 2007).

II.2.4 - Autonomie et complémentation fonctionnelle

a) Les éléments autonomes

Les éléments « autonomes » sont des copies dont la séquence est complète, dont le cadre de lecture n'est pas interrompu par un codon stop et qui portent tous les signaux de régulation : site d'initiation, de polyadénylation, motifs permettant l'excision et l'insertion... Ces éléments, par exemple les *LINEs* ou les Hélitrons, utilisent souvent la machinerie enzymatique de l'hôte pour synthétiser les nouvelles copies. Mais les éléments autonomes sont capables de synthétiser tous les autres produits protéiques nécessaires à leur cycle de transposition.

b) Les éléments non-autonomes

Il semble que la majorité des copies présentes dans les génomes ne soient pas autonomes. L'apparition, dans leur région codante, d'un codon stop ou d'un frameshift, ou encore la délétion d'une région nécessaire au cycle de transposition, leur a fait perdre leur capacité à synthétiser leur propre machinerie enzymatique, et donc leur capacité à transposer de manière autonome. Qu'il s'agisse de mutations bloquant la synthèse de la protéine ou conduisant à la production d'une polyprotéine inactive, ces évènements peuvent provoquer la perte d'autonomie de la copie mutée, qui devient alors incapable d'accomplir seule son cycle de transposition. De nombreux éléments non-autonomes ont été décrits, comme les éléments *Morgane* du blé qui ne codent vraisemblablement pour aucune polyprotéine (Sabot and Schulman, 2006), le rétrotransposon *Vine-1* dont la séquence codante est dépourvue de domaine RT-RnaseH (voir page 53) (Verries *et al.*, 2000), les éléments *Spip* et *Squiq* respectivement dérivés des éléments *RIRE3* et *RIRE8* du riz par des délétions dans leurs régions internes (Vitte *et al.*, 2007a), ainsi que d'autres éléments LARD (Kalendar *et al.*, 2004), SINE (Kajikawa and Okada, 2002), TRIM (Witte *et al.*, 2001; Antonius-Klemola *et al.*, 2006) et MITE (Bureau and Wessler, 1994).

c) Mobilité des éléments non-autonomes par complémentation fonctionnelle

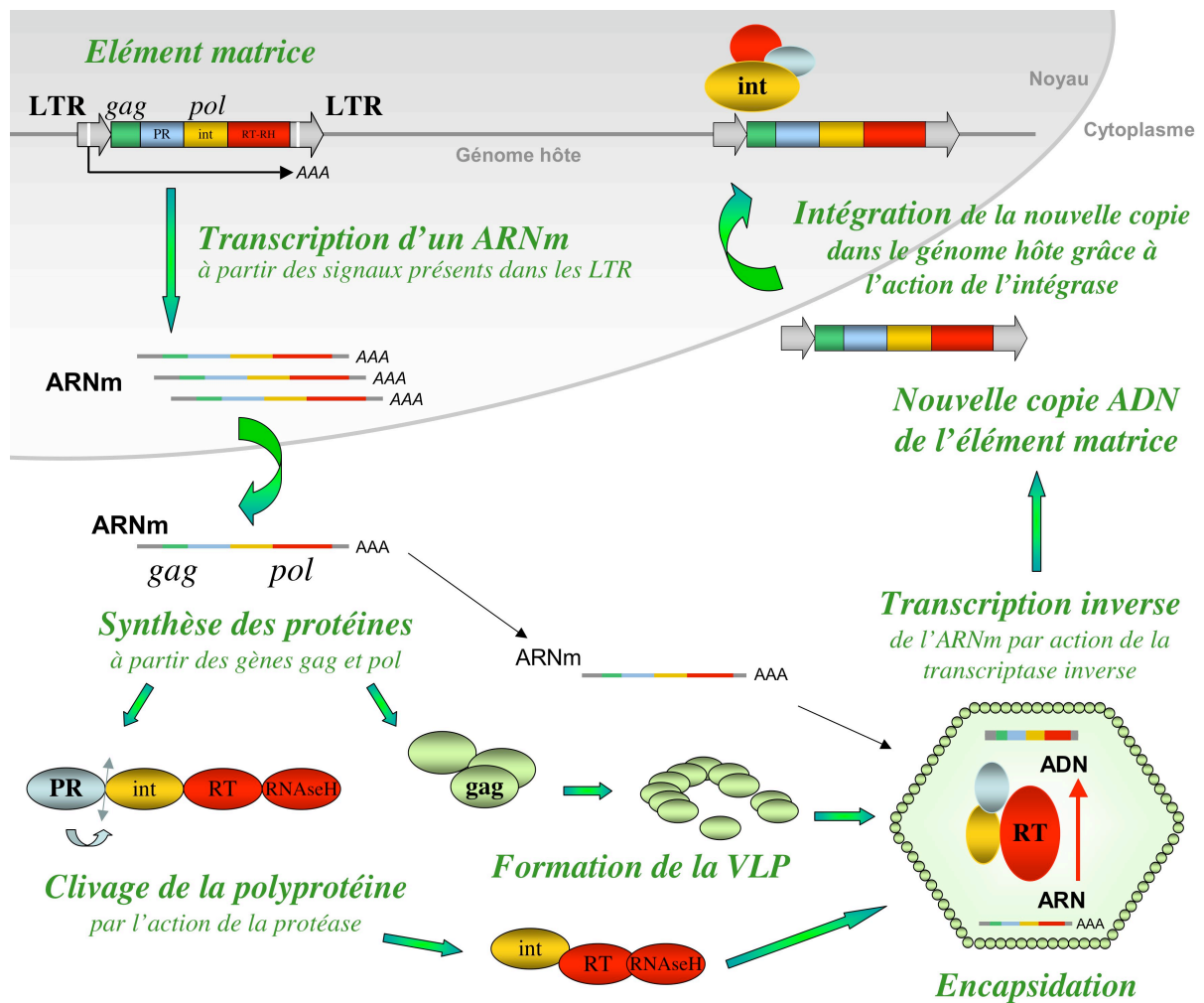
Les copies non-autonomes ne sont pourtant pas systématiquement immobilisées. De manière surprenante, l'élément *Dasheng*, un LARD, s'est multiplié dans le génome du riz au cours des 500000 dernières années, alors qu'il ne contient pas de séquence codante suffisamment

longue pour synthétiser les protéines nécessaires à sa mobilité (Jiang *et al.*, 2002a; Jiang *et al.*, 2002b). Cet élément présente des domaines PPT (Poly-Purine Tract) et PBS (Primer-Binding Site) très similaires à ceux des éléments *Retrosat1/RIRE2* (Ohtsubo *et al.*, 1999). Ces deux domaines, qui jouent un rôle important dans le cycle de rétrotransposition (voir page 31), pourraient favoriser le recrutement par *Dasheng* des protéines codées par des copies *Retrosat1/RIRE2* entières et fonctionnelles, et ainsi permettre l'accomplissement du cycle de l'élément non-autonome (Hu *et al.*, 1995).

Dans certains cas, l'activation en *trans* d'éléments non-autonomes par des copies actives rend donc possible leur multiplication dans le génome. D'autres exemples de complémentation fonctionnelle ont été décrits chez les transposons non-autonomes, comme l'élément *dSpm*, qui a besoin de l'apport en *trans* des protéines codées par les éléments autonomes *Spm* (Gierl, 1996). Un phénomène de « *cross-hybridization* » (hybridation croisée) permet l'amplification de l'élément miniature *mPing* chez le riz. Cet élément, dont la structure est proche de l'élément autonome *Ping*, semble pourtant lié à l'activité d'un élément plus éloigné, l'élément *Pong*. Les deux transposons se seraient donc « renvoyé la balle », le premier donnant naissance aux copies délétées, et le second fournissant ensuite la transposase nécessaire à la multiplication des copies délétées. De la même façon, il semble que les éléments LINEs mobilisent les éléments SINEs, bien que ces derniers ne dérivent pas des LINEs (Kajikawa and Okada, 2002).

Les éléments MITE sont également mobilisés en *trans* par des transposases codées par d'autres transposons (Bureau and Wessler, 1994). Dans le modèle de Feschotte et collaborateurs, les éléments MITE dérivent de transposons complets (Feschotte *et al.*, 2002). L'amplification, par la mobilisation en *cis* ou en *trans*, d'un petit nombre d'éléments tronqués permet la formation d'une (ou plusieurs) « sous-famille(s) » d'éléments MITE homogènes. La transposition d'un élément autonome a probablement conduit à la formation de copies plus ou moins délétées, qui se seraient ensuite amplifiées en très grand nombre, allant jusqu'à environ 100000 copies chez le riz.

Des exemples de parasitisme entre éléments mobiles ont également été décrits chez des virus dont les particules virales ne sont pas fonctionnelles, mais qui utilisent celles de virus compétents (Lazzarini *et al.*, 1981; Ebner and O'Callaghan, 2006).



LTR = Long Terminal Repeats (contenant les signaux d'initiation et de terminaison de la transcription). PR = Protéase.
 int = intégrase. RT = Transcriptase Reverse. RH = RNaseH. AAA = queue poly(A).

Figure 15 : Cycle de rétrotransposition d'un rétrotransposon à LTR.

d) Transcription et mobilité

Il est important de noter que la transcription d'un élément ne signifie pas pour autant qu'il est capable de réaliser son cycle de transposition complet. Il est en effet très difficile d'établir un lien direct entre la quantité d'ARN transcrit et le nombre de nouvelles insertions apparues *in fine*.

II.2.5 - Le cycle de rétrotransposition des rétrotransposons à LTR

L'ARN transcrit à partir d'un rétrotransposon à LTR joue deux rôles dans le cycle de rétrotransposition (Figure 15). Dans un premier temps, il permet la synthèse des protéines nécessaires au cycle de rétrotransposition : la protéine GAG (Group Associated Antigen), qui compose la VLP (Virus-Like Particule) et la protéine POL qui peut être considérée comme la machinerie enzymatique. Le second rôle joué par l'ARN, une fois encapsidé dans la VLP, est de servir de matrice à la transcription inverse des nouvelles copies ADN de l'élément qui pourront ensuite s'intégrer dans le génome.

a) Transcription d'un ARN messenger

La séquence du rétrotransposon à LTR est transcrite à partir de signaux présents dans les LTR. La transcription de l'ARN débute généralement au niveau de la TATA box située en amont de la région R de la LTR 5'. La synthèse de l'ARN se poursuit jusqu'à la région R de la LTR-3' et le site de poly-adénylation, incluant donc la région interne qui code pour la polyprotéine (Figure 16). Les rétrotransposons utilisent la machinerie cellulaire de l'hôte pour leur transcription, puisqu'ils utilisent le mécanisme classique de transcription par la polymérase II pour synthétiser leurs ARNm.

b) Traduction

Comme pour les ARNm cellulaires, les ARNm du rétrotransposon sont transférés dans le cytoplasme et, lorsqu'il s'agit d'éléments autonomes, l'ARN y est traduit en une polyprotéine correspondant aux régions *gag* et *pol*.

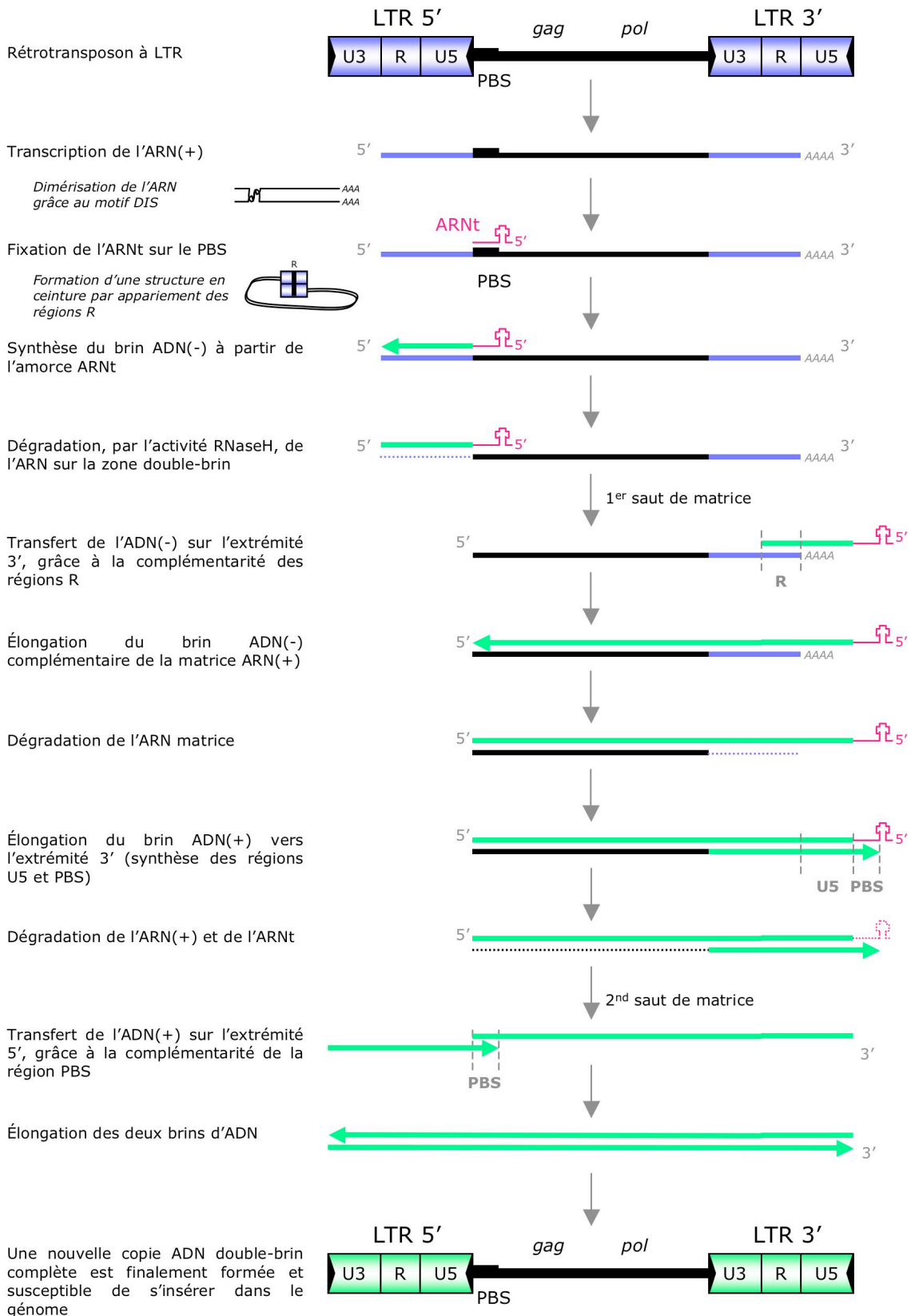
Dans le cas de *BARE-1*, la polyprotéine est synthétisée à partir d'un seul cadre de lecture, puis clivée en sous-unités fonctionnelles GAG et POL (Sabot and Schulman, 2006). Il s'agit du

mécanisme le plus fréquemment observé chez les rétrotransposons et qui semble caractériser le groupe des Ty1-*copia*. A l'inverse, les éléments du groupe Ty3-*gypsy* contiennent majoritairement deux cadres de lecture ouverts *gag* et *pol* distincts.

Chez les rétrotransposons, la synthèse protéique débute donc généralement en amont de la région *gag*. Dans le cas de l'élément *Bs1* du maïs la traduction *in vitro* produit deux protéines : la première issue de l'ORF1, et la seconde suite à un décalage du cadre de lecture (frameshift) qui permet la traduction simultanée de l'ORF2 et aboutit à la synthèse d'une seule protéine plus grande (Jin and Bennetzen, 1989). Dans d'autres cas, un second site interne d'initiation de la traduction, moins efficace, peut parfois permettre l'ancrage du ribosome directement au niveau de la séquence *pol* du rétrotransposon (Meignin *et al.*, 2003). Ce procédé peut avoir deux avantages pour la réussite du cycle de rétrotransposition. Tout d'abord, ce site donne un accès interne au ribosome et peut permettre aux copies portant un codon stop dans la région *gag* de leur séquence codante de produire indépendamment la protéine POL. Ensuite, le site d'initiation de la traduction présent en amont de la *gag* est probablement plus efficace que le site interne donnant accès à la région *pol*. De ce fait, la production de protéines GAG est sensiblement plus importante que la production de protéines POL. Sachant que la particule virale du virus HIV humain est par exemple composée de plus de 5000 protéines structurales GAG (Briggs *et al.*, 2004), l'expression différentielle des régions *gag* et *pol* peut permettre de répondre à un besoin plus important en GAG pour l'assemblage de la particule VLP.

c) Encapsidation

Les particules VLP, aussi appelées « nucléocapsides », sont formées par l'assemblage des protéines GAG (Figure 15). La protéine GAG des rétrotransposons contient trois domaines fonctionnels : le domaine Capside, qui permet la polymérisation des protéines pour former la VLP ; le domaine Nucléocapside, qui favorise les interactions avec les acides nucléiques grâce à des structures en doigts de zinc et des résidus basiques ; et le domaine Matrice, qui permet l'association avec la protéine de l'enveloppe (lorsqu'elle existe). Chez les rétrovirus, ces trois domaines sont présents chez un précurseur de la protéine GAG puis sont individualisés par clivage.



LTR = Long Terminal Repeat, composée des régions U3, R et U5. PBS = Primer Binding Site, site de liaison de l'ARNt. ARNt = ARN de transfert.

Figure 16 : Synthèse de la nouvelle copie du rétrotransposon (étapes de transcription et transcription inverse).

L'ARN correspondant au rétrotransposon est incorporé dans la VLP par un mécanisme d'encapsidation (packaging) lors duquel l'ARN qui a servi à synthétiser la GAG est sélectionné (Figure 15). Chez les rétrovirus, la sélection des ARN se fait via un motif *PSI* (packaging signal) localisé après le PBS, et qui est reconnu par le domaine Nucléocapside (Dalton *et al.*, 2001; Evans *et al.*, 2004). Lorsque ce motif est placé en amont d'un gène rapporteur, les ARNm de ce gène sont encapsidés dans la nucléocapside du rétrovirus correspondant. Si aucun motif *PSI* n'a encore été identifié chez les rétrotransposons, il a été supposé qu'un mécanisme similaire permet la sélection des ARN à encapsider chez les rétrotransposons (Sabot and Schulman, 2006).

d) Synthèse de la nouvelle copie ADN du rétrotransposon

Dans la VLP, l'ARN sert de matrice à la synthèse d'une nouvelle copie ADN de l'élément (Figure 15). Plusieurs études tendent à prouver que l'ARN des rétrotransposons se présente alors sous une forme dimérique, comme c'est le cas chez les rétrovirus. Un signal de dimérisation, formant des boucles symétriques, semble permettre la reconnaissance et l'interaction spécifique de deux molécules d'ARN (Paillart *et al.*, 2004).

L'ARN forme une sorte de « ceinture », fermée par la jonction des régions R des LTR-5' et 3'. La transcription inverse est réalisée par la transcriptase inverse qui se sert de l'ARNm comme matrice (Figure 16). Elle est initiée au niveau du PBS par la fixation d'un ARN de transfert, qui porte en général un acide aminé méthionine. La transcription inverse du brin d'ADN(-) se poursuit vers la LTR-5' et s'arrête après la région R. Une fois ce fragment du brin (-) d'ADNc synthétisé, la RNaseH dégrade la matrice d'ARN correspondante. L'ADNc ainsi libéré est transféré sur l'extrémité 3'-R du second brin d'ARN matrice qui compose la ceinture, c'est le premier saut de matrice. La transcription inverse du brin (-) se poursuit en partant cette fois de la LTR-3' et en direction de la séquence interne, incluant le domaine PPT, puis la région interne (ORF puis UTL), et finalement le PBS. A cette étape, l'ADNc n'est encore que simple-brin et n'inclut pas la séquence de la LTR-5'. La RNaseH dégrade la séquence ARN matrice, cette fois au niveau du motif PPT et de la LTR-3'. Le brin (+) est ensuite synthétisé, tout d'abord au niveau de la LTR-3' puis, après un second saut de matrice du brin (+) de l'ADNc, reprend au niveau de la LTR-5' et inclut la totalité de la séquence de l'élément. Le brin ADN (-) sert à son tour de matrice pour la synthèse du brin (+).



Le rétrotransposon

Il est composé par la région interne et les LTR 5' (à gauche) et 3' (à droite). Les LTR sont bordées de séquences répétées inversées (triangles noirs) généralement composées des dinucléotides TG-CA, et sont constitués de trois régions (U3, R et U5) qui contiennent les signaux d'initiation et de terminaison de la transcription. Des motifs promoteurs et le signal de poly-adénylation sont présents dans la région U3. Les LTR 5' et 3' d'une copie peuvent présenter des séquences sensiblement différentes (représentées par les deux couleurs).



L'ARN messenger

Il est issu de la transcription du rétrotransposon. Il commence à la fin de la région U3 de la LTR 5' et se termine à la fin de la région R de la LTR 3'. La présence des deux extrémités répétées va permettre le premier saut de matrice lors du cycle de répllication.



La copie néo-synthétisée

A la fin du cycle de rétrotransposition, la copie nouvellement insérée dans le génome présente normalement deux LTR aux séquences identiques. Par la suite, leurs séquences peuvent évoluer et accumuler indépendamment des mutations. Il est donc possible, en se basant sur la divergence des LTR, de « dater » l'insertion de la copie. De chaque côté de l'élément, les duplications de la séquence de l'hôte ne sont *a priori* pas les mêmes que celles de la copie d'origine, ce qui permet de distinguer les copies d'une famille.

Figure 17 : Conséquences du mécanisme de rétrotransposition sur la structure de la copie néo-synthétisée.

Par ce mécanisme, une molécule d'ADN complémentaire double-brin, composée de deux LTR identiques, est donc synthétisée à partir de la copie ARNm matrice (Figure 16) (Wilhelm *et al.*, 1999; Wilhelm and Wilhelm, 2001; Le Grice, 2003).

e) Intégration dans le génome hôte

La copie d'ADN néosynthétisée dans la nucléocapside est transférée dans le noyau pour s'insérer dans le génome hôte (Figure 15). Ce mécanisme et le rôle exact de la VLP dans le transfert de l'ADNc vers le noyau restent à déterminer.

L'intégrase se fixe sur des motifs situés aux extrémités des LTR. Elle crée une cassure asymétrique dans l'ADN génomique, au niveau d'un site cible, qui est généralement composé de quelques paires de bases. La copie ADNc du rétrotransposon s'insère au niveau de la cassure, qui est ensuite comblée par le système endogène de réparation de l'ADN génomique. Cette dernière étape a pour conséquence la duplication du site cible de chaque côté de la copie insérée. Cette duplication, appelée Target Site Duplication (TSD), constitue en quelque sorte la « signature » de la nouvelle copie intégrée. Il semble que l'insertion d'une nouvelle copie dans le génome soit conditionnée par la nature hétérochromatique du site ciblé, ainsi que par la présence ou non d'autres éléments à proximité.

II.2.6 - Principe de l'estimation de l'âge des copies (datation)

A la suite du cycle de rétrotransposition, le rétrotransposon à LTR nouvellement inséré dans le génome est encadré par des LTR 5' et 3' qui présentent des séquences théoriquement identiques (Figure 17). Au cours du temps, les deux séquences LTR vont ensuite accumuler des mutations de manière indépendante. La divergence des LTR 5' et 3' est, en théorie, proportionnelle au nombre d'années écoulées depuis l'insertion du rétrotransposon. Autrement dit, les deux LTR divergent d'autant plus que la copie est ancienne. Cette divergence est utilisée pour estimer la date d'insertion de copies dans le génome car elle peut être convertie en une durée à partir d'une estimation du taux de substitution des LTRs : $T = D / 2t$ (avec T , le temps écoulé depuis l'insertion, D le taux de divergence des deux LTR estimé à partir d'un modèle d'évolution comme Kimura à 2 paramètres, et t le taux de substitution de la séquence étudiée). Cette méthode a été utilisée à l'origine pour retracer la dynamique

d'amplification d'éléments insérés dans la région *Adh* chez le maïs (SanMiguel *et al.*, 1998). Elle est depuis couramment utilisée pour dater, en millions d'années, les insertions de rétrotransposons à LTR chez les plantes (Wright and Voytas, 1998; Devos *et al.*, 2002; Jiang *et al.*, 2002a; Jiang *et al.*, 2002b; SanMiguel *et al.*, 2002; Laten *et al.*, 2003; Gao *et al.*, 2004). Vitte et collaborateurs ont proposé de corriger cette estimation en utilisant la correction de Jukes et Cantor (1969). Ainsi, la durée depuis l'insertion est estimée par la formule : $D = 3/4 \times \ln(1 - 4/3d)$, avec d la divergence observée entre les LTR par le modèle de Jukes et Cantor (1969). Cette méthode a, par exemple, permis d'estimer que les familles *Spip* et *Squiq* se sont principalement amplifiées au cours des deux derniers millions d'années chez le riz (Vitte *et al.*, 2007a).

II.2.7 - La notion de « famille » chez les rétrotransposons

a) Erreurs lors de la transcription inverse

La rétrotransposition est un processus peu fidèle car l'étape de transcription inverse introduit régulièrement des mutations dans la séquence des copies néoformées. En effet, la transcriptase inverse ne possède pas de fonction correctrice, ce qui peut aboutir à un taux de $2,5 \times 10^{-5}$ erreurs par nucléotide et par cycle, un taux 100 à 1000 fois supérieur à ceux des ADN polymérases cellulaires (Gabriel *et al.*, 1996; Preston, 1996; Boutabout *et al.*, 2001). De plus, l'activité de transcription inverse n'est pas spécifique d'un ARN particulier. Si les deux molécules encapsidées dans la VLP sont généralement identiques car issues du même rétrotransposon, le saut de matrice sera sans conséquence sur la structure de la nouvelle copie. Mais il peut arriver que deux ARN différents soient encapsidés dans la même particule. Un saut de matrice peut alors conduire à la formation de structures complexes (Sabot and Schulman, 2007), comme pour les éléments *BARE2* chez l'orge (Vicient *et al.*, 2005) et *VejuL* chez le blé (Sabot *et al.*, 2005), ou comme nous le verrons pour deux éléments de la famille *Tvv1* chez la vigne (Moisy *et al.*, 2007, voir page 127). Ainsi, les nouvelles copies diffèrent souvent de la copie maître d'origine.

b) Divergence des copies

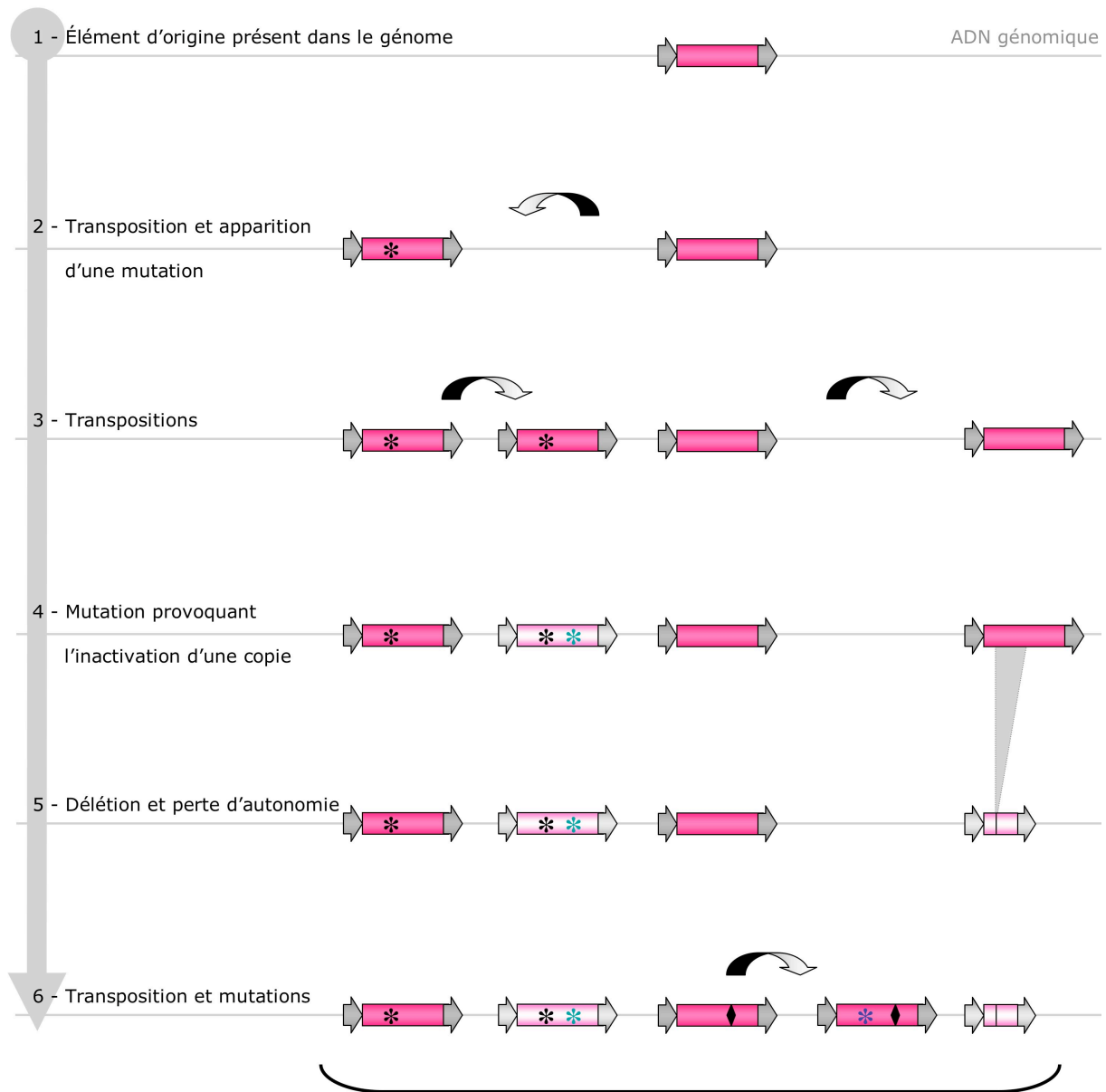
Une fois intégrées dans le génome, les copies peuvent encore diverger au cours du temps par l'accumulation de mutations (mutations ponctuelles, délétions, insertions...) et par des

phénomènes de recombinaison, pouvant conduire à la formation de nouvelles structures. La recombinaison homologue inégale entre les deux LTR d'un élément complet peut par exemple conduire à la formation d'un seul LTR recombiné et flanqué du site cible dupliqué autour de l'élément complet de départ, un « solo-LTR » (Shirasu *et al.*, 2000). Chez le riz, les trois rétrotransposons *hopi*, *Retrosat1* et *RIRE3* comptent au total 75 solo-LTR pour 114 copies complètes. La famille *RIRE3* présente même 25 solo-LTRs pour seulement 4 copies complètes (Vitte and Panaud, 2003). Au sein d'une famille de rétrotransposons, les copies solo-LTR sont donc parfois plus nombreuses que les copies complètes. Leur proportion dépend probablement de l'âge de la famille et de la taille des LTR (Vitte and Panaud, 2003) mais varie également d'une espèce à l'autre. Le génome du maïs contient principalement des copies intactes et peu de solo-LTR (SanMiguel *et al.*, 1996; Devos *et al.*, 2002). Chez *A. thaliana*, l'analyse de 291 rétrotransposons à LTR, appartenant à 4 familles *Ty1-copia*, 6 *Ty3-gypsy* et 2 familles inconnues, a montré que 87 (29,9%) d'entre eux étaient des copies déléetées, 98 (33,7%) des copies incomplètes et 101 (34,7%) des solo-LTR (Devos *et al.*, 2002). Chez l'orge, les LTR de l'élément *BARE-1* sont 16 fois plus nombreuses que sa région interne, alors que seulement deux fois plus de LTR devraient être détectées si les copies étaient intactes (Vicent *et al.*, 1999). Cet excès de LTR a été attribué à l'abondance de solo-LTR (Shirasu *et al.*, 2000).

Pour toutes ces raisons, une « famille » de rétrotransposons est constituée d'une population de copies présentant des séquences sensiblement différentes, mais toutes apparentées au même élément (Figure 18) (Konieczny *et al.*, 1991; Casacuberta *et al.*, 1995; Marillonnet and Wessler, 1998).

c) Evolution des domaines fonctionnels chez les rétroéléments

La pression de sélection qui s'exerce au niveau de la séquence des éléments, dans le cadre du maintien de fonctions essentielles comme la transcription inverse, engendre une hétérogénéité dans la conservation intra- et inter-familles des domaines fonctionnels. Ainsi, le domaine RT montre un taux d'évolution relativement plus lent que les autres domaines des rétrotransposons. Ce domaine a été utilisé pour calculer les distances évolutives entre les familles de rétrotransposons potentiellement distantes (Flavell, 1986; Xiong and Eickbush, 1990; Dixit *et al.*, 2006). Bowen et McDonald proposent que les éléments peuvent être regroupés au sein de la même famille lorsque leurs séquences en acides aminés présentent une



Différentes copies ayant divergé par l'accumulation de mutations au cours de l'évolution, mais appartenant à la même « **famille** ». Les différentes séquences présentées par ces copies et les direct repeats qui les encadrent permettent de les distinguer.

Figure 18 : Evènements successifs pouvant conduire à la différenciation des copies d'une même famille.

identité supérieure à 90% entre les motifs I à VII du domaine RT, et regroupés en super-famille lorsque ce taux est supérieur à 75% (Bowen and McDonald, 1999).

Le domaine *gag* est relativement spécifique de la famille d'éléments qui la synthétise et cette région évolue jusqu'à deux fois plus vite que la région RT (McClure *et al.*, 1988).

L'analyse des séquences LTR, qui évoluent encore plus rapidement, permet de dater l'insertion des copies (voir page 34) et de caractériser les sous-structures phylogéniques au sein des familles et super-familles de rétrotransposons à LTR, comme les sous-familles de *Retrolyc1* chez *Lycopersicum* (de Araujo *et al.*, 2001), les éléments de la famille *Tvv1* chez la vigne (Pelsy and Merdinoglu, 2002), et les éléments de la super-famille *Tnt1* chez les Solanacées (Manetti *et al.*, 2007).

II.3 - Contrôle de l'activité des éléments transposables

L'ubiquité des éléments transposables dans les génomes contraste avec les effets négatifs pouvant résulter de leur mobilité, suggérant qu'ils ont développé une stratégie leur permettant de co-exister dans le génome de l'hôte, voire de lui être bénéfique, ou que l'hôte contrôle leur activité par différents mécanismes (Figure 19).

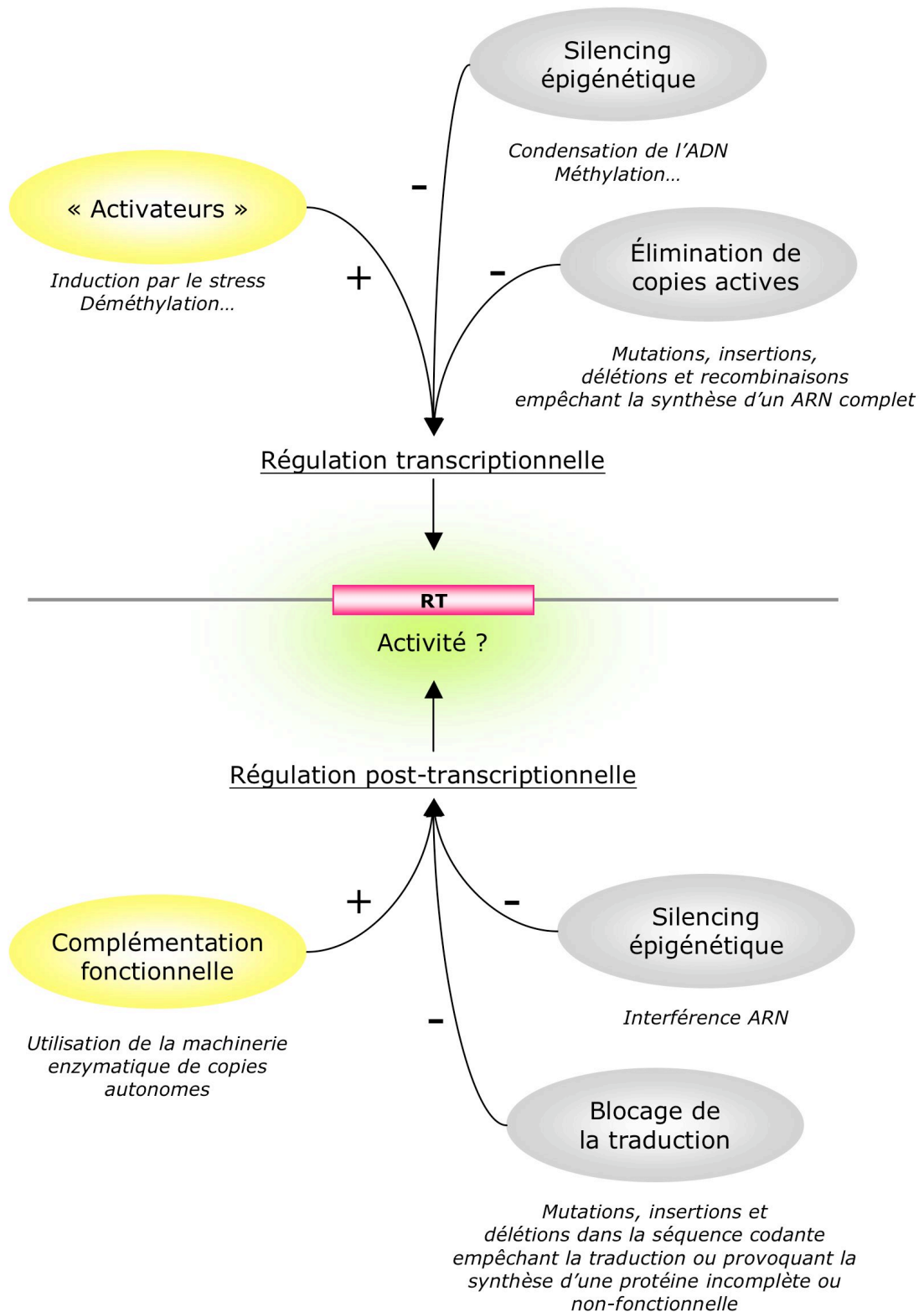
II.3.1 - La sélection : élimination de copies

Plusieurs modèles proposent que la sélection agit contre les effets négatifs provoqués par l'insertion d'un élément dans ou à côté des gènes (Biemont *et al.*, 1997). Les éléments transposables sont susceptibles de s'accumuler préférentiellement dans les régions peu denses en gènes car le risque de perturber un gène y est plus faible. La sélection peut également agir contre les échanges ectopiques en contre-sélectionnant les remaniements chromosomiques résultant de la recombinaison illégitime de copies présentes dans le génome (voir page 47) (Charlesworth *et al.*, 1994).

II.3.2 - Régulation de l'activité des éléments transposables

L'accumulation de mutations, de délétions et les recombinaisons intra- ou inter-éléments sont des mécanismes qui participent de manière importante à l'élimination, à l'inactivation, ou à la perte d'autonomie des éléments transposables. Ces mécanismes contribuent également à la variation de la taille des génomes (voir page 47). D'autre part, la transposition des éléments mobiles peut potentiellement être régulée à chaque étape du cycle et le contrôle de la quantité de transcrits en est le premier niveau. D'après Melayah et collaborateurs, la transcription est l'étape clé du contrôle de l'amplification des éléments *Tnt1* dans le génome du tabac (Melayah *et al.*, 2001) et les facteurs d'induction de l'activité des rétrotransposons sont généralement spécifiques de chaque famille (Vernhettes *et al.*, 1998).

Le niveau d'expression des rétrotransposons à LTR est généralement plus faible que celui des gènes de l'hôte (Wessler *et al.*, 1995; Jaaskelainen *et al.*, 1999). Des séquences EST



RT = Rétrotransposon. + = Effet positif (activation, induction). - = Effet négatif (inhibition, repression, suppression)

Figure 19 : Principaux facteurs susceptibles de réguler l'activité des éléments transposables.

(Expressed Sequence tag) présentant des similarités avec les LTR et les régions internes de rétrotransposons ont été recherchées parmi 780000 EST disponibles pour les plantes. Cette étude a révélé que les éléments transposables représentent 1,2 séquences EST sur 1000 et que cette proportion varie énormément d'une espèce à l'autre. En effet, les EST de rétrotransposons représentent 0,85‰ (12) des EST disponibles pour *Solanum*, 0,98‰ (61) pour le riz, 1,25‰ (153) pour le soja, 2,49‰ (184) pour le maïs, et jusqu'à 23,8‰ (3) pour le tabac (Vicient *et al.*, 2001). D'après Boroluzzi et collaborateurs, les transcrits peuvent être considérés comme abondants lorsque les séquences EST correspondantes représentent plus de 3,6‰ des EST disponibles, modérés entre 3,6‰ et 1,25‰, et rares en dessous de 1,2‰ (Bortoluzzi *et al.*, 2000). Selon ces critères, les rétrotransposons sont modérément exprimés chez 3 dicotylédones sur les 9 étudiés, et chez 4 monocotylédones sur les 6 étudiés (Vicient *et al.*, 2001). Ce constat peut s'expliquer par le fait que les promoteurs présents dans les LTR sont peu efficaces ou parce que ces éléments sont soumis à des régulations transcriptionnelles et post-transcriptionnelles, comme des mécanismes de méthylation de l'ADN, de formation d'hétérochromatine ou d'interférence ARN (Okamoto and Hirochika, 2001). Certains éléments, pourtant complets et intacts, sont maintenus inactifs dans le génome. Si une copie inactivée épigénétiquement conserve la possibilité de se mobiliser, un environnement chromatique répressif ne lui permet pas de produire les protéines nécessaires à son cycle de transposition.

a) Régulation au cours du développement

Les transcrits de rétrotransposons ont été recherchés par RT-PCR ou Northern dans différents organes de plantes. Les transcrits de *Tnt1* ont été recherchés par Northern dans les racines, les tiges, les pétioles, les feuilles, les sépales, les pétales, les étamines et les ovaires de tabac, mais n'ont été détectés que dans les racines et en très petite quantité (Pouteau *et al.*, 1991). Les éléments *Tto1*, *Tos10* et *Tos17* ne sont pas exprimés dans les feuilles et les éléments *Tto1* du tabac et *Tos17* du riz sont transcrits uniquement dans les cellules en culture (Hirochika, 1993; Hirochika *et al.*, 1996). Chez le maïs, les éléments *Opie*, *Huck*, *Cinful* sont exprimés dans les feuilles (Grandbastien, 1998) alors que l'élément *CIRE1* de l'oranger est exprimé dans les racines, mais pas dans les feuilles (Rico-Cabanas and Martinez-Izquierdo, 2007). L'élément *PREM-2* du maïs s'exprime lors des premiers stades du développement des microspores, mais ne s'exprime ni dans le pollen, ni dans les autres tissus (Turcich *et al.*, 1996). Ainsi, l'activité transcriptionnelle de la majorité des rétrotransposons étudiés jusqu'à

présent n'est pas constitutive et semble au contraire étroitement régulée en conditions « normales » de croissance, c'est-à-dire sans stress apparent, puisque ces éléments présentent un profil d'expression tissu-spécifique au cours du développement de la plante.

A l'inverse, quelques familles de rétrotransposons semblent s'exprimer dans plusieurs conditions ou organes. Des transcrits de l'élément *Ogre* ont par exemple été détectés par RT-PCR dans les racines, les feuilles et les fleurs du pois (Neumann *et al.*, 2003). De plus, une étude combinant des approches de Northern, de microarrays et de transformations de plantes a démontré que les éléments transposables de la canne à sucre s'expriment essentiellement dans les cals, mais aussi dans les feuilles, les fleurs et les méristèmes apicaux. Cependant, cette activité transcriptionnelle n'est pas suivie par l'apparition de nouvelles insertions dans le génome de la plante (de Araujo *et al.*, 2005).

Les séquences EST correspondant spécifiquement à 38 rétrotransposons connus (*Ty1-copia* et *Ty3-gypsy*) ont été recherchées dans les banques d'EST obtenues à partir de différents organes : pour 8 d'entre eux, aucune EST n'a été détectée et les autres montraient des similarités avec 1 à 34 EST (Vicient *et al.*, 2001). Au total, environ 26,5% des EST de ces rétrotransposons ont été détectés dans des banques EST construites à partir de feuilles, 2% à partir de racines et les pousses, 30% à partir de fleurs et d'autres organes de l'appareil reproducteur, 11,6% à partir d'embryons et de graines, et enfin 20% à partir de graines (d'après Vicient *et al.*, 2001).

b) Activation des éléments mobiles en conditions de stress

D'après la définition du stress de Madlung et Comai, celui-ci se définit par ses effets négatifs sur la croissance et le développement de l'individu (Madlung and Comai, 2004). Les stress internes, comme les mutations spontanées de gènes ou les divisions cellulaires aberrantes, peuvent avoir des effets sur le métabolisme et la régulation des gènes. Les stress externes sont de nature biotique ou abiotique. Les stress biotiques peuvent être provoqués par des attaques de pathogènes (bactéries, champignons, insectes...), par la consommation de la plante par des herbivores ou encore la compétition entre organismes. Les stress abiotiques résultent de blessures, de conditions environnementales défavorables comme des températures anormales, des carences en eau, en nutriments, les irradiations ou le manque de lumière... Quelque soit leur origine, les stress exercent une pression évolutive très forte sur l'organisme, qui doit alors développer des tolérances, des résistances ou des mécanismes de contournement.

Lors de la remise de son prix Nobel, Barbara McClintock a présenté quatre exemples de stress susceptibles de provoquer de grandes restructurations génomiques via l'activation des éléments transposables ou la modification de la structure des chromosomes : la culture de tissus, l'attaque de pathogènes, les croisements inter-spécifiques et la séparation des lignées cellulaires lors des premiers stades du développement des tissus somatiques (McClintock, 1984).

La culture de tissus

Pendant la culture de tissus, des cellules différenciées sont « réorientées » pour donner des cellules totipotentes, puis « reprogrammées » pour créer un nouveau tissu ou une nouvelle plante. D'après McClintock, cette reprogrammation cellulaire « inflige à la cellule une série d'expériences traumatisantes » (McClintock, 1984). L'activité des éléments transposables dans ces conditions a été mise en évidence pour la première fois chez le tabac, lorsque des transcrits des éléments *Tto1*, *Tto2* et *Tto3* ont été amplifiés dans des ADN complémentaires préparés à partir de protoplastes (Hirochika, 1993). Suite à leur transcription, le nombre de copies de *Tto2* et *Tnt1* est plus important dans les cellules différenciées du nouveau tissu. Dans le cas de l'élément *Tto1*, le nombre de copies est même jusqu'à 10 fois supérieur au nombre de copies présentes dans la cellule de départ, et ce nombre est également plus important dans les plants de tabac régénérés à partir de ces cellules. Chez le riz, les transcrits des trois éléments *Tos10*, *Tos17* et *Tos19* sont également plus nombreux lors de la culture de tissus, et entre 5 et 30 nouvelles copies de *Tos17* ont été détectées dans les plantes régénérées par ce procédé (Hirochika *et al.*, 1996). Chez l'orge, l'élément *BARE-1* est exprimé dans les cals embryogènes et dans les feuilles régénérées à partir de protoplastes (Suoniemi *et al.*, 1996). L'élément *mPing*, un MITE du génome du riz, est mobilisé lors de la culture d'anthères (Kikuchi *et al.*, 2003). Des transcrits d'éléments SINE ont également été détectés dans la culture de cals de colza (Deragon *et al.*, 1996). Enfin, une stratégie de S-SAP, qui permet de comparer le nombre de copies de rétrotransposons présents dans différents génomes, a permis de mettre en évidence un polymorphisme d'insertion chez les plantes régénérées à partir de cellules de feuilles de tabac. De nouvelles insertions de *Tnt1* ont été détectées dans 25% des plantes régénérées à partir de protoplastes (Melayah *et al.*, 2001).

La culture de tissus provoque la blessure et la digestion des parois cellulaires lors de l'isolement des protoplastes. Des transcrits de l'élément *Ogre* ont été détectés par RT-PCR dans des protoplastes et dans des feuilles de pois dont la surface avait été écorchée à l'aide de

papier de verre (Neumann *et al.*, 2003). L'élément *CIRE1* de l'oranger est sur-exprimé dans les feuilles blessées par des coupures réalisées au scalpel, ou soumises à des hormones telles que le méthyl-jasmonate et l'auxine (Rico-Cabanas and Martinez-Izquierdo, 2007). De la même façon, des transcrits de l'élément *Tto1* ont été détectés dans les feuilles de tabac entre 2 et 48 heures après l'application d'une blessure, l'expression étant plus importante dans les jeunes feuilles que dans les feuilles déjà développées (Takeda *et al.*, 1998).

Le lien direct entre la culture de tissus et l'activation de la mobilité des éléments transposables n'est pas encore établi mais pourrait résulter d'une diminution de la méthylation de l'ADN. Dans des plants de riz régénérés par embryogenèse somatique, l'élément *Tos17* s'est par exemple inséré dans le gène *phytochrome A* et dans un gène de la famille des récepteurs kinase (Hirochika *et al.*, 1996), ce qui suppose que les mutations d'insertions consécutives à la mobilisation des éléments transposables par les stress ne sont pas sans conséquences sur l'organisme.

Attaque de pathogènes

L'attaque de la plante par des pathogènes, comme des champignons, des bactéries ou des virus, est un stress qui affecte régulièrement les plantes au cours de leur vie. Il a été démontré que le rétrotransposon *Tnt1* est activé par les éliciteurs fongiques qui hydrolysent les cellules végétales (Grandbastien *et al.*, 1997; Melayah *et al.*, 2001). De même qu'après la blessure, l'expression de *Tto1* augmente entre 2 et 48h après l'application d'extraits fongiques sur les feuilles de tabac (Takeda *et al.*, 1998). L'expression de *Tto5* augmente après un traitement à l'acide salicylique et après l'inoculation de virus (Grandbastien, 1998). L'infection de plantes par le virus *Barley stripe mosaic virus* (BSMV) induit un choc génomique et des mutations provoquées par la mobilisation des éléments transposables (Dellaporta *et al.*, 1984), notamment de l'élément *Bs1* dont l'insertion dans le gène *Adh* du maïs est consécutive à l'inoculation du virus (Johns *et al.*, 1985).

Présence de motifs régulateurs dans les LTR

L'expression des gènes de la plante en réponse à l'attaque d'un pathogène est généralement induite par une cascade de transduction de signaux, qui déclenche finalement la réponse de l'organisme. Certains éléments transposables sont susceptibles d'être activés par les mêmes signaux. Suoniemi et collaborateurs ont étudié en détail la structure des LTRs de l'élément *BARE-1* en construisant plusieurs systèmes d'expression transitoire et en plaçant différents

fragments de LTR en amont d'un gène rapporteur. Ils ont ainsi pu confirmer que certains motifs présents dans les LTR sont indispensables à la transcription de *BARE-1*, comme la TATA box, et ont également pu caractériser des motifs responsables de la répression ou de la surexpression du rétrotransposon dans les feuilles ou les cals (Suoniemi *et al.*, 1996). Un motif de 31 pb, répété en tandem, a été détecté dans la région U3 des LTRs des éléments actifs *Tnt1A* et *Tto1* et semble impliqué dans l'activation de *Tnt1A* dans les protoplastes et en réponse aux éliciteurs fongiques (Casacuberta *et al.*, 1997; Grandbastien *et al.*, 1997). Un motif de 13 pb permet également la régulation de l'expression de *Tto1* dans les cals embryogènes et en réponse à la blessure, aux éliciteurs fongiques et au méthyl-jasmonate (Takeda *et al.*, 1999). Des motifs présents dans les LTRs de *Tnt1A* et *Tto1* montrent des similarités avec des gènes de défense ou des motifs régulateurs de l'expression de gènes induits par les stress (Grandbastien, 1998).

Des motifs particuliers, présents dans les LTRs, sont donc sensibles aux signaux émis lors de l'attaque de pathogènes et permettent l'induction « directe » de l'expression de certains rétrotransposons par les stress.

Cependant, l'expression des rétrotransposons peut également être une conséquence « indirecte » des effets du stress sur l'organisme. Les éléments transposables, comme les gènes de l'hôte, peuvent être constitutivement « éteints », par exemple par la méthylation des promoteurs, et n'être réactivés que lorsqu'un stress provoque la déméthylation de l'ADN (Madlung and Comai, 2004).

Stress abiotiques et adaptation des génomes via l'activité des éléments transposables

Quelques études décrivent l'activation d'éléments transposables en réponse à des stress abiotiques. Chez la Gueule de loup, *Antirrhinum majus* (aussi appelée snapdragon), la mobilité des éléments *Tam* dépend de la température, et l'élément *Tam3* se transpose jusqu'à 1000 fois plus souvent à 15°C qu'à 25°C (Hashida *et al.*, 2003). L'induction de cet élément par des températures faibles est influencée par sa position dans le génome, alors que son immobilité à température plus forte ne dépend pas de sa position (Kitamura *et al.*, 2001). D'après Hashida et collaborateurs, la mobilité de *Tam3* est corrélée à la méthylation de l'ADN via la sensibilité à la température de protéines telles que les méthyl-transférases (Hashida *et al.*, 2003).

Lorsque les éléments transposables ont été mis en évidence par Barbara McClintock, celle-ci suggéra qu'ils pouvaient être activés en réponse à une sollicitation du génome par des facteurs environnementaux (McClintock, 1984). Du point de vue évolutif, l'activation des éléments transposables génère une part de la diversité génétique qui peut favoriser l'adaptation de l'organisme aux changements de conditions environnementales.

Un tel phénomène a peut-être eu lieu chez l'orge sauvage. En Israël, le polymorphisme d'insertion de l'élément *BARE-1* permet de distinguer très clairement deux populations, pourtant distantes de 300 mètres au maximum, poussant sur les versants Nord et Sud d'un canyon (Kalendar *et al.*, 2000). Le nombre de copies de *BARE-1* est en effet nettement plus important dans les plantes trouvées en altitude et en condition aride, au sommet du versant Sud. Cette adaptation différentielle est probablement le résultat d'une pression de sélection qui s'est exercée soit sur la taille du génome, soit sur les effets physiologiques consécutifs à la mobilité de *BARE-1*. La présence, dans la séquence promotrice des LTR de *BARE-1*, d'un motif de réponse à l'acide abscissique, caractéristique des gènes induits par le stress hydrique, pourrait expliquer la corrélation entre l'induction de *BARE-1* et les conditions de sécheresse (Kalendar *et al.*, 2000).

Chez le riz, l'amplification du nombre de copies du transposon *mPing* a été plus importante dans les variétés tropicales et sub-tropicales qui se sont adaptées aux climats tempérés (Jiang *et al.*, 2003b). De plus, l'élément *mPing* s'est inséré dans le gène *CONSTANS*, un homologue chez le riz du gène contrôlant la date de floraison, ce qui pourrait avoir participé à l'adaptation des variétés de riz aux températures plus froides et aux saisons de culture plus courtes. D'après les auteurs de cette étude, la mobilisation des éléments *mPing* en réponse au stress provoqué par changement d'environnement aurait participé à la diversification de ces variétés.

Finalement, la polypléidisation des génomes entraîne des changements dans la méthylation de l'ADN et dans la régulation de l'expression de gènes, des facteurs qui pourraient favoriser la mobilisation des éléments transposables. Si aucune relation directe entre ce phénomène et la transposition des éléments transposables n'a été décrite, il semble que la polypléidisation puisse affecter la transcription des éléments transposables. Une étude sur le maïs ($2n = 2.500\text{Mb}$) a en effet montré que sur 407000 séquences EST, 0,014% montraient des similarités avec des rétrotransposons (Meyers *et al.*, 2001) alors que les éléments

transposables représentent 2,4% des EST chez *Triticum aestivum* (6n = 17.000Mb) (Li *et al.*, 2004) et 2,3% chez la canne à sucre (8n = 3.150Mb) (Vettore *et al.*, 2003).

II.3.3 - Interférence ARN

De petits ARN peuvent initier le silencing au niveau post-transcriptionnel ou conduire à des modifications de la structure chromatinienne, et potentiellement modifier l'expression des éléments transposables (Figure 19). Le processus par lequel la production des ARN d'éléments transposables seraient produits n'est pas connu, mais ces éléments possèdent des caractéristiques structurales susceptibles d'en faire des cibles de ce mécanisme de régulation (revue par Slotkin *et al.* 2007).

II.4 - Participation des éléments transposables à la structure et à la dynamique des génomes

II.4.1 - Impact sur la taille du génome hôte

Le paradoxe de la valeur C, qui résulte du constat que la taille d'un génome n'est pas corrélée à la complexité biologique de l'organisme (Thomas, 1971), peut s'expliquer chez les plantes par la proportion de régions non-géniques et leur influence sur les variations de la taille des génomes (Bennetzen *et al.*, 1998). L'amplification des éléments transposables et les recombinaisons qu'ils peuvent provoquer participent à ce phénomène.

a) Invasion du génome

En dépit des mécanismes limitant leur transposition et donc leur nombre de copies, plusieurs études montrent que les éléments transposables sont parfois présents en un très grand nombre de copies dans les génomes. Des expansions massives peuvent être tolérées durant l'évolution et la taille d'un génome peut alors être sensiblement influencée par la présence et la multiplication des éléments transposables (Figure 20) (Kidwell, 2002). Si la multiplication de petits éléments n'influence pas toujours de manière significative la taille du génome, l'amplification d'éléments de grande taille peut jouer un rôle d'une toute autre ampleur. Par exemple, l'élément *Ogre*, dont certaines copies dépassent 22kb, avec des LTR de plus de 5kb, occupe environ 5% du génome du pois (Neumann *et al.*, 2003). Les 13700 copies de l'élément *BARE-1* représentent 2,9% du génome de l'orge (Vicient *et al.*, 1999). Les 2200 copies complètes de l'élément *CIRE1* représentent 2,9% du génome de *Citrus sinensis* (Rico-Cabanas and Martinez-Izquierdo, 2007). Il a été démontré que le génome de *Oryza australiensis*, une espèce sauvage de riz, a été la cible de l'insertion de plus de 90000 copies d'éléments transposables au cours des 3 derniers millions d'années, entraînant le doublement de la taille du génome (Piegu *et al.*, 2006). Les trois rétrotransposons *Kangourou*, *RIRE1* et *Wallabi* représentent près de 60% du génome de *O. australiensis*. La dynamique d'amplification et d'élimination des copies de 41 familles de rétrotransposons a ensuite été évaluée chez le riz cultivé *O. sativa* (Vitte *et al.*, 2007b). D'après les auteurs de cette étude, la

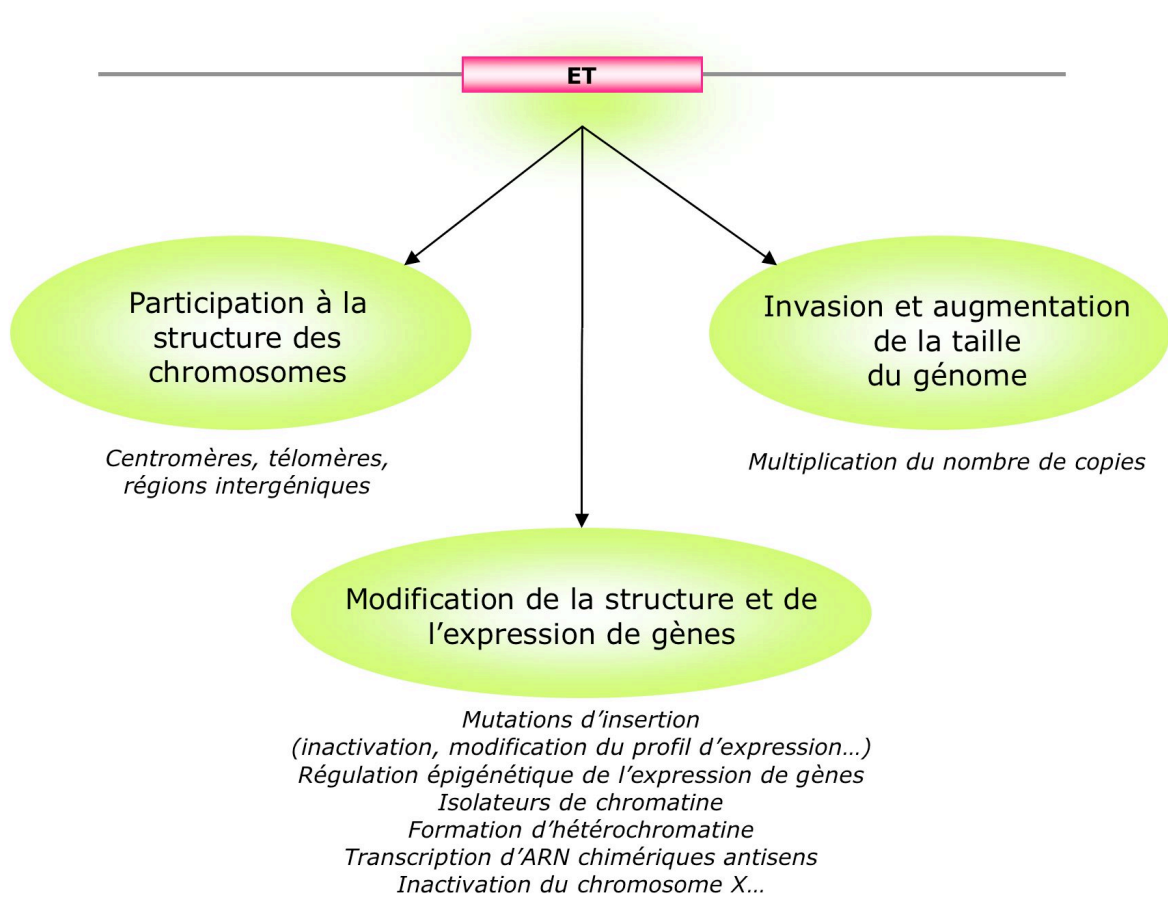


Figure 20 : Conséquences de l'activité et de la présence des éléments transposable sur la structure et la dynamique du génome.

majorité des éléments présents dans le génome du riz résulte d'insertions qui se sont produites au cours des deux derniers millions d'années, chaque famille présentant un profil d'amplification différent dans le temps. Chez la drosophile, la taille des génomes varie d'une espèce à l'autre et est corrélée au nombre d'éléments transposables présents sur les chromosomes (Vieira *et al.*, 2002).

Lorsqu'il s'insère pour la première fois dans un génome « vierge », suite à un transfert horizontal, à un croisement ou à un autre mécanisme, un élément autonome peut se multiplier avant qu'une régulation génétique ou épigénétique ne se mette en place. C'est par exemple le cas des éléments *P*, *hobo* et *I* du génome de la drosophile (Biemont and Vieira, 2005). Les éléments transposables donc peuvent envahir très rapidement un génome, comme l'illustre le cas de l'élément *P*, présent dans les espèces de *D. melanogaster* du monde entier alors qu'il est apparu pour la première fois il y a 40 ans (Anxolabehere *et al.*, 1988).

b) Recombinaison et perte de matériel génomique

Plusieurs études décrivent des événements de recombinaison entre éléments mobiles (voir page 35). Chez l'orge, la recombinaison entre les LTRs de deux éléments *BARE-1* différents a provoqué la formation de solo-LTR et l'élimination de fragments génomiques de tailles importantes au locus *rar1* (Shirasu *et al.*, 2000). La formation de solo-LTR représente une force qui participe de manière très importante à la diminution de la taille des génomes (Vitte and Panaud, 2003).

Les délétions qui affectent la séquence interne des éléments transposables représentent également un mécanisme participant à l'élimination d'ADN chez les plantes (Devos *et al.*, 2002). De tels événements ont également été décrits dans les éléments transposables du maïs (Masson *et al.*, 1987; Marillonnet and Wessler, 1998) et dans les LTR de rétrotransposons de blé (Wicker *et al.*, 2001).

II.4.2 - Participation des éléments transposables à la structure du chromosome

a) Les éléments transposables comme constituants des centromères

Chez l'homme, le centromère est constitué de multiples séquences répétées en tandem, entourées d'une région périphérique riche en éléments transposables, essentiellement des rétrotransposons de type LINE et SINE (Schueler and Sullivan, 2006). Chez la drosophile, les répétitions centrales sont très courtes et interrompues par des rétrotransposons spécifiques, comme les éléments *HMS Beagle* et *412*, qui appartiennent à la classe Ty3-gypsy (Sun *et al.*, 2003). Chez les plantes, des rétrotransposons très bien conservés sont présents dans les centromères, ce qui peut impliquer une fonction elle aussi très conservée (Jiang *et al.*, 2003a). Chez *A. thaliana*, la région centrale des centromères est composée de satellites interrompus par des rétrotransposons, alors que la région périphérique porte plutôt des transposons (The Arabidopsis Initiative, 2000). Les éléments transposables ont donc probablement donné naissance aux collections de répétitions satellites présentes dans les régions centromériques.

b) Participation au maintien des télomères

Les extrémités des chromosomes des cellules eucaryotes, les télomères, sont formées de nombreuses répétitions en tandem sur lesquelles sont fixées des protéines spécifiques. Ces répétitions en tandem sont ajoutées à l'extrémité des chromosomes par la télomérase transcriptase inverse, un mécanisme qui présente des similitudes avec l'intégration des rétrotransposons sans LTR. L'hétérochromatine subtélomérique est composée d'éléments transposables complets et tronqués. Chez la drosophile, les rétrotransposons *HeT-A* et *TART* s'insèrent spécifiquement dans les extrémités chromosomiques et remplacent les séquences répétées généralement observées dans ces régions (Levis *et al.*, 1993). Ainsi, ces éléments transposables limitent probablement le raccourcissement progressif des télomères au cours des divisions cellulaires, un phénomène qui entraîne une importante instabilité chromosomique et qui est directement impliqué dans le processus du vieillissement. Leur insertion s'effectue de manière orientée et régulée au cours du développement (Casacuberta and Pardue, 2005; Pardue *et al.*, 2005) par des mécanismes de contrôle, notamment par l'interférence ARN (Casacuberta and Pardue, 2006).

Classe	ET détectés dans les exons de 2060 gènes		ET détectés dans la séquence codante de 439 gènes	
	Nb copies	Taille (pb)	Nb copies	Taille (pb)
Eléments de classe I	710	205 073	269	65 111
LINE	8	1 556	0	0
LTR <i>Ty1-copia</i>	103	37 256	52	22 790
LTR <i>Ty3-gypsy</i>	290	69 531	81	7 812
Autres	309	96 730	136	34 509
Eléments de classe II	1 005	130 068	166	21 517
Autres éléments	1 714	171 348	144	5 522
Total	3 429	506 489	579	92150

Tableau 3 : Eléments transposables détectés dans les exons et dans la séquence codante de gènes du riz (d'après Sakai *et al.*, 2007).

II.4.3 - Influence sur l'expression des gènes de l'hôte

a) Insertion d'éléments transposables dans les gènes

Les données accumulées lors de l'analyse du génome du riz tendent à démontrer que les éléments transposables ont joué un rôle dans l'évolution de cette espèce. Plus de 10% des gènes structuraux du riz contiennent des éléments transposables. Cependant, ces derniers sont très peu représentés dans les exons, puisque seulement 2% des gènes contiennent des éléments transposables dans leur régions exoniques ou codantes (Tableau 3). Ce résultat suggère que la pression de sélection peut avoir joué un rôle important dans la protection de ces séquences contre l'insertion d'éléments transposables (Sakai *et al.*, 2007). Chez l'homme, des éléments transposables sont présents dans environ 4% des gènes (Nekrutenko and Li, 2001). Chez la Gueule de loup, l'insertion des éléments *Tam1* et *Tam3* dans les promoteurs et de *Tam2* dans la séquence codante des gènes codant pour la couleur des fleurs a provoqué l'apparition de phénotypes bariolés (Coen *et al.*, 1986). Aussi, l'insertion et la conservation de ces éléments dans les séquences codantes est potentiellement une force évolutive vers l'acquisition de nouvelles fonctions biologiques (Figure 21).

Les éléments transposables contribuent de manière importante aux régions régulatrices de nombreux gènes. Dans le génome humain, 25% des gènes sont localisés à proximité d'un élément transposable susceptible d'en moduler l'expression, et pour 3% des gènes humains annotés, les rétrotransposons apportent les signaux de régulation de la transcription (Van de Lagemaat *et al.*, 2003). Chez *Arabidopsis*, un élément transposable inséré dans l'intron du gène FLC (Flowering Locus C) provoque la régulation de l'expression du gène en dirigeant des modifications de la chromatine répressive par la voie de petits ARN interférents. Ces petits ARN sont générés à partir d'autres copies de l'élément transposable présentes à des endroits différents du génome (Liu *et al.*, 2004).

b) Régulation épigénétique de l'expression de gènes de l'hôte

En soumettant des gènes proches à la régulation épigénétique qui les cible, les éléments transposables peuvent produire de la variabilité dépendant de modifications épigénétiques et non pas de changements nucléotidiques (Figure 21). Les épiallèles, qui sont des changements épigénétiques héréditaires mais réversibles de l'expression d'un gène, peuvent ainsi donner une

expression différente dans des cellules génétiquement identiques. Ces allèles sont parfois sous le contrôle d'un élément transposable ou d'une répétition en tandem présent à proximité, et donnent des phénotypes contrastés comme des effets mosaïques (Slotkin and Martienssen, 2007).

Transcrits antisens

Plusieurs exemples de participation de transcrits d'éléments transposables à la répression ou la sur-expression de gènes adjacents ont été décrits. Chez la souris, les promoteurs présents aux extrémités du rétrotransposon *IAP* peuvent être à l'origine de la production d'un ARN antisens qui entraîne la transcription de la séquence adjacente à son site d'insertion, et celle du gène *Agouti* qui contrôle la couleur du pelage. La quantité de transcrits du gène *Agouti*, ainsi que la couleur du pelage de la souris, sont héréditaires et dépendent de l'état épigénétique du rétrotransposon (Morgan *et al.*, 1999).

Chez l'homme, des transcrits synthétisés à partir du promoteur antisens de l'élément *LINE1* sont dirigés vers l'extérieur de l'élément et participent à l'expression de gènes voisins dans des tissus spécifiques (Nigumann *et al.*, 2002). Chez le maïs, l'activité de la famille de transposon *Mutator* régule deux épiallèles. Dans les régions où l'élément *Mutator* est actif, des phénotypes mutants pour les deux épiallèles apparaissent, caractérisés par la présence de tâches et de points de nécrose sur les feuilles. La transcription de l'épiallèle *hcf106* est initiée par l'élément non-autonome *Mu1* uniquement lorsque l'élément autonome présent ailleurs dans le génome est dans un état inactif (Martienssen and Baron, 1994).

Les promoteurs présents dans la LTR-5' sont également présents dans la LTR-3' et il peut arriver qu'un élément tronqué ou incomplet ait perdu les signaux de terminaison de la transcription. Ainsi, la transcription d'un ARN peut parfois s'étendre au delà de la fin de l'élément et inclure un gène adjacent (Figure 21). Les transcrits qui en résultent pourront alors être soumis aux contrôles transcriptionnels et post-transcriptionnels qui affectent normalement l'élément transposable. Chez le blé, les transcrits de l'élément *Wis2-1A* participent à la régulation des gènes adjacents (Kashkush *et al.*, 2003).

Formation d'hétérochromatine facultative

La propagation de l'hétérochromatine vers des gènes adjacents peut aboutir à l'apparition de phénotypes différents, associés au silencing de nouveaux gènes (Figure 21). Ce phénomène est connu sous le nom de Position Effect Variegation (PEV), et apparaît comme une variation

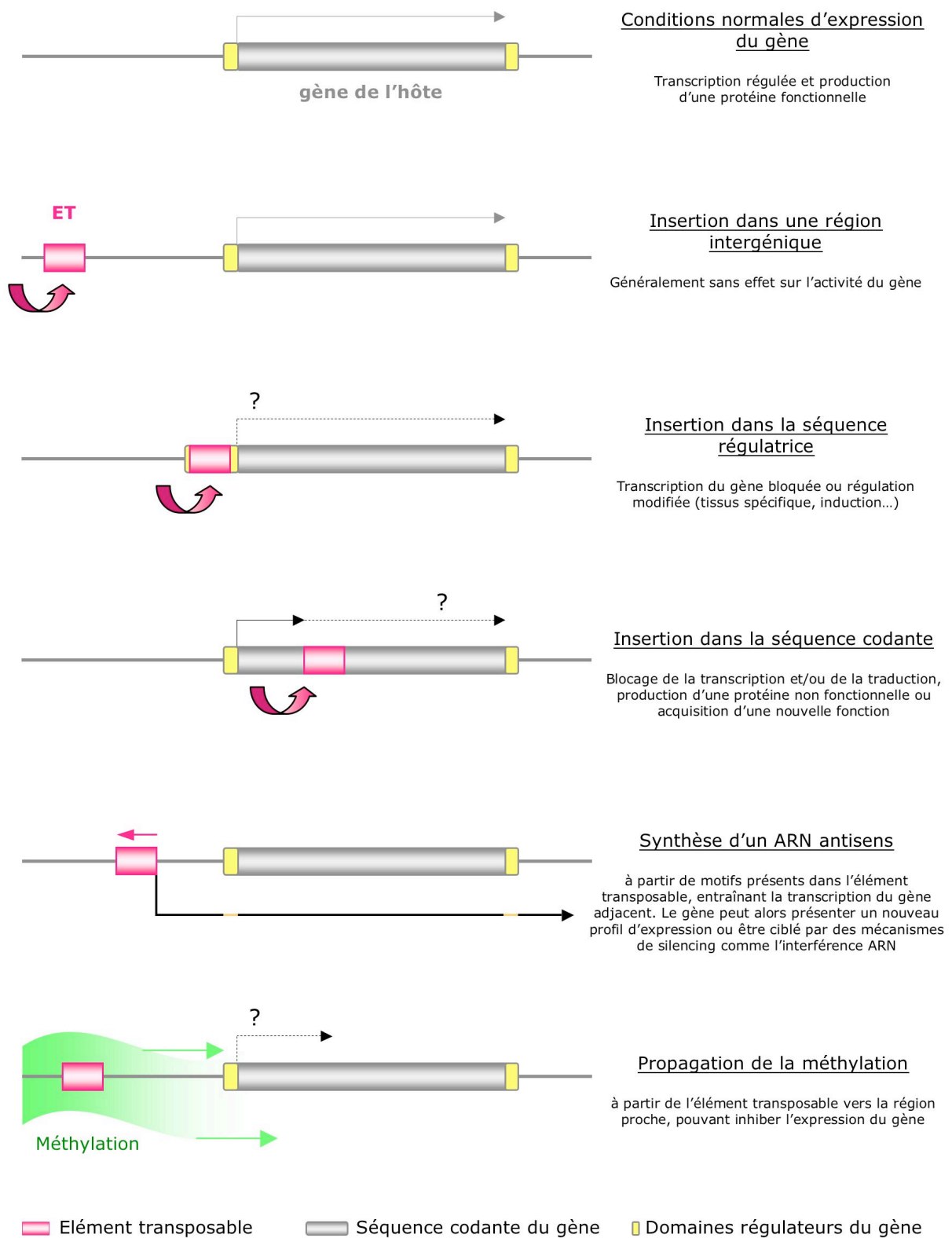


Figure 21 : Conséquences possibles de l'insertion d'un élément transposable dans ou à proximité d'un gène.

de l'expression en fonction de la position du gène sur le chromosome. L'insertion d'un transgène dans l'hétérochromatine peut donner ce type de phénotype, parfois influencé par la proximité de l'insertion avec des éléments transposables. Chez *D. melanogaster*, l'insertion du transgène *white* dans le génome donne normalement un phénotype présentant des yeux rouges. Lorsqu'il s'insère dans des régions « actives », comme dans un gène, le transgène s'exprime correctement. A l'inverse, l'insertion du transgène dans ou à proximité d'un élément transposable donne une expression variable, aboutissant à l'observation d'un phénotype bariolé (Sun *et al.*, 2004). Les auteurs de cette étude supposent que les éléments transposables sont parfois des centres pour la formation d'hétérochromatine de type « facultative », qui peut s'étendre linéairement sur le chromosome sur plus de 10kb.

Des isolateurs de chromatine

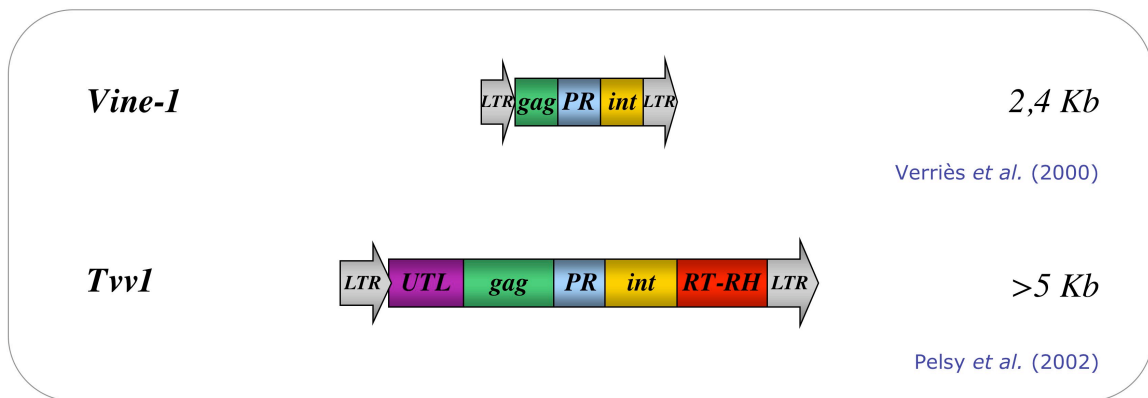
Différentes études ont permis d'identifier des séquences, appelées des « *isolateurs* », qui séparent des zones de chromatine active et inactive, et semblent agir sur la transcription, soit comme de simples barrières bloquant la progression de protéines régulatrices le long de la séquence, soit en modifiant l'organisation ou la structure des fibres chromatiniennes (Bell *et al.*, 2001). Il a été démontré que les éléments transposables peuvent parfois jouer ce rôle. Par exemple, l'isolateur *gypsy*, une séquence de 350 pb découverte dans le rétrotransposon *gypsy* de *D. melanogaster*, est à l'origine de réorganisations spatiales de la chromatine qui sont corrélées avec l'état transcriptionnel de la cellule (Gerasimova *et al.*, 2000). L'isolateur *gypsy* forme, avec plusieurs protéines, un complexe qui rattache des fibres de chromatine à la membrane nucléaire. Cette réorganisation de la chromatine forme des boucles qui isolent ou rapprochent certaines séquences du génome, et peuvent bloquer les interactions de divers promoteurs avec leurs enhanceurs. Les isolateurs participent ainsi au contrôle de l'expression des gènes en limitant les interactions entre les promoteurs et leurs éléments régulateurs, à travers la modulation de l'organisation spatiale de la chromatine (Cai and Levine, 1997).

Participation des éléments LINE à l'inactivation du chromosome X

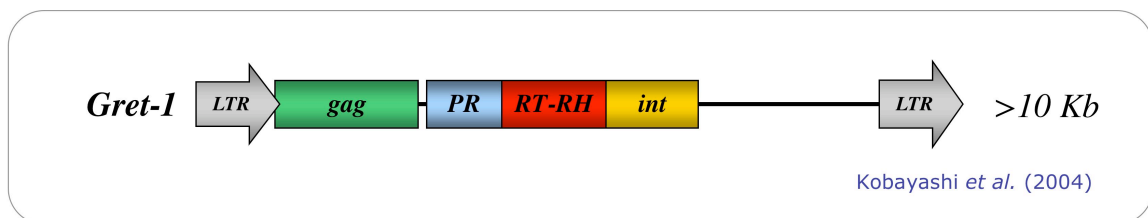
Chez les mammifères, pendant le développement embryonnaire des femelles, un des deux chromosomes X passe d'un état euchromatique actif à un état hétérochromatique inactif. Ce processus est connu sous le nom d'« *inactivation du chromosome X* ». Il aboutit à l'extinction transcriptionnelle de plusieurs milliers de gènes et permet d'assurer un dosage génétique équivalent entre les sexes, la femelle compensant la double dose de gènes du chromosome X qu'elle possède par rapport au génome XY des mâles (Lyon, 2006). De manière surprenante,

l'inactivation ne touche qu'un seul des deux chromosomes X, alors que son homologue est lui aussi présent dans le même noyau, mais reste actif. Une fois établie, l'inactivation est stable et transmise au cours des divisions cellulaires. Ce processus est initié en un point précis appelé « centre d'inactivation » puis se propage sur le reste du chromosome. Les chromosomes X de nombreux mammifères, notamment celui de l'homme, sont riches en éléments LINE. De plus, de courtes séquences répétées, présentant une homologie avec les rétrotransposons, produisent des transcrits qui semblent réguler l'inactivation du chromosome X dans des cellules souches embryonnaires différenciées de souris (Cohen *et al.*, 2007). Il a été proposé que les rétrotransposons LINE participent au processus en favorisant l'initiation et la propagation du silencing le long du chromosome (Lyon, 2006).

Ty1-copia-like



Ty3-gypsy-like



LTR = Long Terminal Repeats, PR = protéase, int = intégrase, RT = Transcriptase Reverse, RH = RNaseH

Figure 22 : Représentation schématique de la structure des trois rétrotransposons à LTR décrits chez la vigne.

II.5 - Les rétrotransposons du génome de la vigne

Comme nous l'avons vu tout au long de ce chapitre, les éléments transposables ont été mis en évidence par différentes stratégies. Plusieurs éléments ont été découverts dans la séquence codante de gènes, comme l'élément *Tnt1* dans le gène de la nitrate réductase (Grandbastien *et al.*, 1989) ou *Tst1* dans le gène de la phosphorylase chez la pomme de terre (Camirand and Brisson, 1990). Le criblage de banques d'ADN génomique a permis la détection de l'élément *BARE-1* chez l'orge en utilisant des sondes LTR (Manninen and Schulman, 1993) et de l'élément *PDR1* chez le pois, caractérisé en tant que séquence répétée dispersée dans le génome (Lee *et al.*, 1990). Des expériences de RT-PCR à partir de tissus en culture ont permis d'identifier les rétrotransposons *Tto1-Tto3* chez le tabac (Hirochika, 1993) et *Tos17-Tos20* chez le riz (Hirochika *et al.*, 1996). Aujourd'hui, la généralisation des outils bioinformatiques et le séquençage complet de plusieurs génomes de plantes permettent la recherche de séquences répétées et la caractérisation rapide et à grande échelle de nouvelles familles d'éléments transposables.

Lorsque mon travail de thèse a débuté, en 2005, seules deux copies de rétrotransposons et une famille d'éléments avaient été décrites. Depuis, le séquençage et l'analyse du génome de l'accession PN40024 a permis d'estimer le nombre d'éléments transposables présents dans le génome de la vigne (Jaillon *et al.*, 2007).

II.5.1 - L'élément *Vine-1*

Le premier élément transposable identifié chez la vigne, *Vine-1*, a été trouvé inséré dans le quatrième exon du gène *Adhr* (Verries *et al.*, 2000). Ce rétrotransposon à LTR du groupe des Ty1-*copia*-like a une séquence de 2392 pb encadrée par deux LTRs de 287 pb, identiques à 99,3%. Cependant, la copie *Vine-1* est incomplète, car tronquée dans la partie 3' de sa séquence codante : elle présente un seul cadre de lecture ouvert, qui code potentiellement pour une protéine de 581 acides aminés, correspondant à la région *gag* et aux domaines protéase et intégrase de la région *pol*, mais dépourvue de domaine RT-RNaseH (Figure 22). Ainsi, ce rétrotransposon n'est vraisemblablement pas autonome.

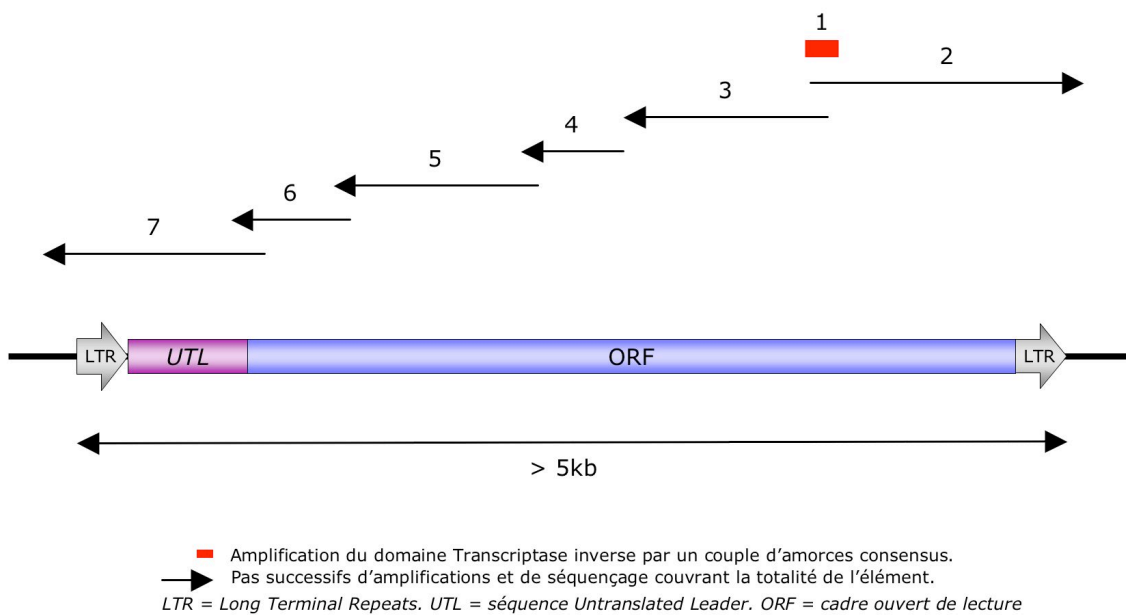


Figure 23 : Caractérisation de la famille *Tvv1* par marche sur chromosome (d'après Pelsy *et al.*, 2002).

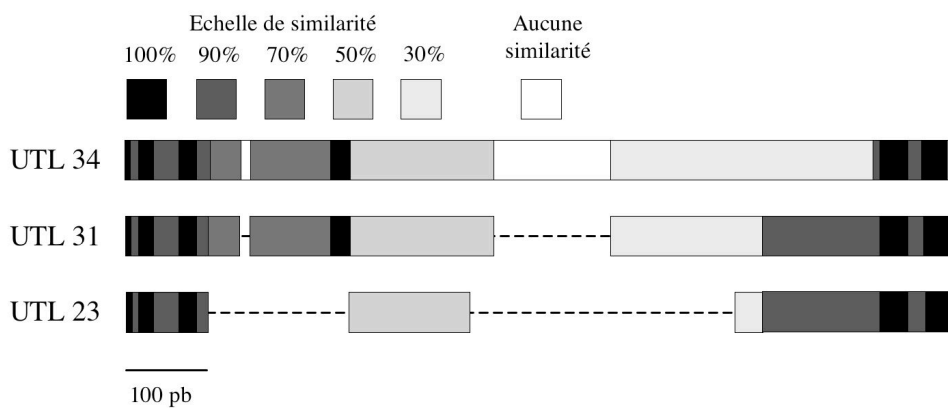


Figure 24 : Comparaison de la structure des séquences UTL de trois éléments de la famille *Tvv1* (d'après Pelsy, 2007).

L'élément *Vine-1* est présent au même site d'insertion *Adhr* chez différentes variétés de *V. vinifera*, mais absent chez les autres espèces de *Vitis* analysées, ce qui suppose que cette copie s'est intégrée dans le génome après la séparation des *Vitis*, mais avant la distinction des variétés de l'espèce *V. vinifera* (Verries *et al.*, 2000). Utilisé comme marqueur S-SAP, *Vine-1* présente plus de polymorphisme que les marqueurs AFLP (Amplified Fragment Length Polymorphism) (16% contre 0%) et permet la distinction des clones de la variété Savagnin (Labra *et al.*, 2004).

II.5.2 - La famille Tvv1

Une famille de rétrotransposons de type Ty1-*copia*, baptisée *Tvv1* pour Transposon of *V. vinifera*, a été reconstituée par une stratégie de marche sur le chromosome (Pelsy and Merdinoglu, 2002). A partir d'une première séquence amplifiée grâce à des amorces dégénérées ciblant les motifs I et III de la transcriptase inverse, des étapes d'amplifications successives en 5' et en 3' ont permis de reconstruire la séquence d'une copie consensus (Figure 23). Les séquences successives n'appartenant pas toutes à la même copie, cette stratégie ne permet pas caractériser la séquence d'une copie complète *sensus stricto*. Cependant, cette stratégie a permis de définir les caractéristiques structurales de la famille *Tvv1*, c'est-à-dire une séquence comprise entre 4970 et 5550 pb, bordée de LTRs de 149 à 157 pb (Pelsy and Merdinoglu, 2002). Nous avons depuis validé ces résultats par l'identification d'une copie *Tvv1* complète dans un clone BAC de la variété Vidal blanc, et la caractérisation de toutes les copies *Tvv1* présentes dans le génome de PN40024 (voir Partie IV, Moisy *et al.*, 2007).

Une région UTL très variable

La taille de la séquence UTL (aussi appelée UTR, Untranslated Leader Region), c'est-à-dire région localisée entre la LTR-5' et la région codante (Figure 22), est très variable d'une copie à l'autre au sein de la famille *Tvv1*. Cette variabilité est essentiellement due à de grandes insertions ou délétions qui touchent la partie centrale de la séquence l'UTL, alors que ses extrémités 5' (proche de la LTR) et 3' (vers l'ORF) sont plus conservées entre les copies *Tvv1* (Figure 24) (Pelsy, 2007).

En raison de cette particularité, l'amplification du domaine UTL par PCR donne un profil multi-bandes sur ADN qui a permis d'estimer que le génome de la variété Riesling contient

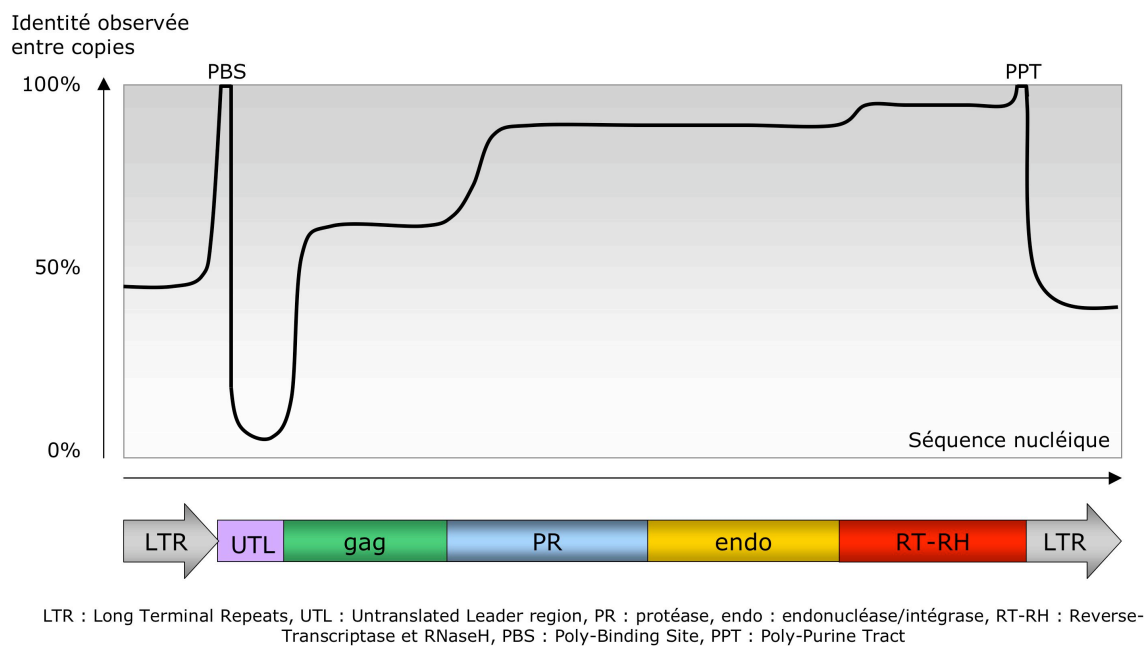


Figure 25 : Représentation schématique du pourcentage d'identité nucléique moyen entre les copies de la famille *Tvv1*, selon la région considérée.

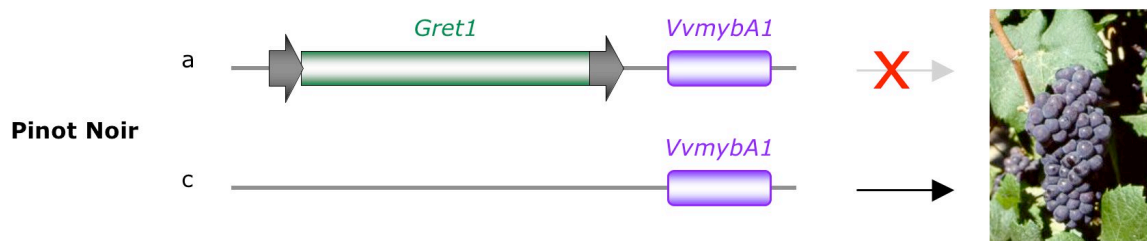
au moins 28 copies *Tvv1* (Pelsy and Merdinoglu, 2002). Cette étude a révélé que les différents domaines qui composent la séquence *Tvv1* présentent des degrés de conservation très variables (Figure 25). De plus, l'amplification de ce domaine sur plusieurs accessions de *Vitaceae* donne des profils spécifiques qui sont conservés entre les clones d'une même variété (Pelsy, 2007). Si l'organisation de la diversité génétique des Vitacées au niveau inter- et intra-variétale déduite des profils d'amplification UTL est comparable à celle obtenue par l'utilisation de marqueurs microsatellites, les profils UTL permettent de distinguer plus précisément les vignes américaines et asiatiques (Pelsy, 2007). Ce nouveau type de marqueur, appelé RUP (Retrotransposons UTL Polymorphism) est comparable à un « code-barre ADN » stable entre les clones (Pelsy, 2007).

II.5.3 - L'élément Gret1

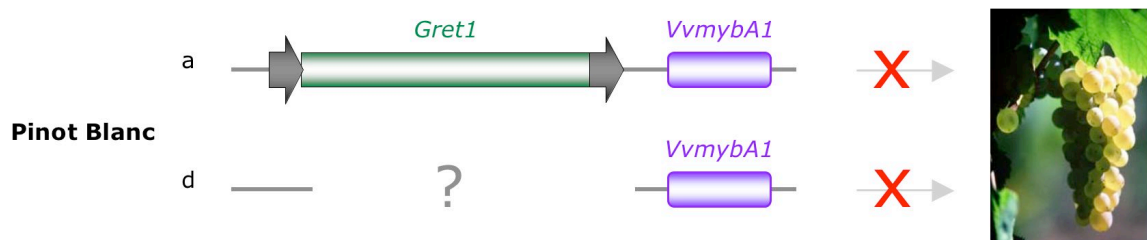
Le premier rétrotransposon complet du génome de la vigne, *Gret1* (pour Grape retrotransposon 1), a été décrit en 2004 comme étant responsable de la mutation du gène *VvmybA1*, qui code pour un facteur de transcription de la voie de biosynthèse des anthocyanes (Kobayashi *et al.*, 2004). Ces auteurs montrent que l'insertion de *Gret1*, entre le gène et son promoteur empêche la synthèse du facteur de transcription et celle des anthocyanes qui donnent normalement la couleur à la pellicule de baie de raisin (Figure 26). L'élément *Gret1* est un rétrotransposon à LTR de type Ty3-*gypsy*, dont la séquence, encadrée par deux LTR de 824 pb chacune, fait au total une taille de 10422 pb (Figure 22).

II.5.4 - Les éléments transposables du génome de PN40024

En 2001, la comparaison des séquences de clones BAC de Cabernet sauvignon avec les séquences présentes dans les bases de données Swiss-PROT a permis d'estimer qu'environ 41% d'entre elles correspondent à des éléments transposables (Tomkins *et al.*, 2001). Cette proportion est relativement importante en comparaison du génome du riz qui a une taille similaire à celui de la vigne. Le génome du riz (430 Mb), par exemple, est constitué à environ 30% d'éléments transposables (International Rice Genome Sequencing Project, 2005) et le génome d'*A. thaliana* (145 Mb) n'en présente « que » 10% (The Arabidopsis Initiative, 2000). Ce résultat a été confirmé par l'analyse de la séquence complète du génome de



Dans l'allèle a, la présence de Gret1 en amont du gène bloque la synthèse des anthocyanes.
L'allèle c permet la synthèse des anthocyanes qui donnent la couleur à la pellicule de baie



La structure de l'allèle d n'est pas déterminée, mais pourrait correspondre à une délétion. Chez le Pinot blanc, aucun allèle ne permet la synthèse des anthocyanes. Les pellicules de baies sont blanches.

Figure 26 : Structure génomique des clones de Pinot Noir et de Pinot Blanc au locus *VvmybA1* (d'après Yakushiji *et al.*, 2006).

PN40024 publiée en 2007 (Jaillon *et al.*, 2007). La détection *in silico*, par BlastX et par annotations manuelles, a permis de montrer la présence de toutes les classes et superfamilles d'éléments transposables dans le génome de la vigne (Jaillon *et al.*, 2007). Dans cette étude, 111 876 éléments transposables ont été annotés manuellement. Les résultats présentés dans le Tableau 4 montrent une très large prédominance des éléments de classe I sur les éléments de classe II et les Hélitrons. Le nombre d'éléments de classe I est supérieur à 105000 alors que moins de 6500 éléments de classe II ont été mis en évidence. Les éléments de classe I représentent également une part nettement plus importante du génome de *V. vinifera* : 17%, contre 0,4% pour les éléments de classe II.

L'analyse de la distribution des séquences répétées sur les fractions de génome annotées a révélé que les introns sont des régions relativement riches en répétitions et en éléments transposables, puisque 12,4% d'entre elles contiennent des transposons (Jaillon *et al.*, 2007). Selon ces auteurs, 75% des éléments transposables présents dans les séquences introniques correspondent vraisemblablement à des éléments de types LINEs. Ces éléments semblent donc contribuer de manière importante à la taille des introns. La densité d'éléments transposables sur les chromosomes présente cependant un profil relativement complémentaire à celui de la densité de gènes, puisque les régions riches en gènes sont pauvres en éléments transposables, et inversement (Jaillon *et al.*, 2007).

Type	Famille	Nombre copies	Couverture (Kb)	% génome	
Protéines détectées par BlastX	Eléments de classe I		33 118	50863,3*	10,65
	Non-LTRs	LINEs	5 504	6 792,9	1,42
	LTR	<i>Ty1-copia</i>	17 293	24 640,8	5,16
		<i>Ty3-gypsy</i>	9 632	17 659,6	3,70
		Autres	88	103,6	0,02
	Autres		601	166,6	0,35
	Eléments de classe II		1 797	1 975,9	0,41
	Hélitrons		109	58,9	0,01
	Total		35 024	52 898,0	11,08
	Annotation manuelle	Eléments de classe I		105 532	81 363,7
Non-LTRs		LINEs	15 216	12 131,1	2,54
LTR		<i>Ty1-copia</i>	56 890	39 848,3	8,35
		<i>Ty3-gypsy</i>	14 093	15 339,8	3,21
		Autres	18 688	13 191,5	2,76
Autres			645	853,0	0,18
Eléments de classe II		6 344	2 040,9	0,43	
Hélitrons			0	nc	nc
Total		111 876	83 404,7	17,47	

* D'après les données présentées dans Jaillon et al. 2007, la somme des éléments de classe I représenterait en fait 49 364Kb, soit une taille moyenne de 1,491kb et environ 9,87% du génome.

Tableau 4 : Estimation du nombre d'éléments transposables présents dans le génome de *V. vinifera* (d'après Jaillon et al. 2007).

Chapitre III : Présentation des travaux de thèse

Propagée par multiplication végétative sur des pas de temps très longs, la vigne peut être considérée comme un modèle pour l'étude des plantes clonales qui évoluent essentiellement par l'accumulation de mutations somatiques. Dans un contexte général de compréhension des mécanismes biologiques qui participent à l'évolution des génomes au cours des cycles de multiplication végétative, l'objectif de la thèse est d'étudier la structure et la dynamique de transcription des rétrotransposons du génome de la vigne.

Pour cela, le projet de recherche a été développé en trois parties, répondant aux questions suivantes :

1) Peut-on identifier de nouvelles familles dans le génome de la vigne ?

2) Quel est le niveau de transcription des différentes familles de rétrotransposons, au cours du cycle de développement d'une part, et en réponse à l'induction par les stress biotiques et abiotiques d'autre part ?

3) Quel sont le niveau et l'origine de diversité des éléments de la famille *Tvv1* ?

Les réponses que nous apportons à ces questions sont présentées dans les trois prochaines parties de ce manuscrit. Elles doivent nous apporter des indications sur la diversité et la mobilité des familles de rétrotransposons, afin d'évaluer dans quelle mesure ces éléments sont susceptibles de générer de la diversité par l'insertion de nouvelles copies et, ainsi, de participer à l'adaptation de la vigne au cours son évolution.

La partie II de ce manuscrit présente, sous la forme d'un article en cours de préparation, la caractérisation de dix nouvelles familles de rétrotransposons du génome de la vigne, ainsi que les données supplémentaires acquises sur cinq autres éléments décrits dans les bases de données.

La partie III regroupe les résultats de l'étude, par des stratégies complémentaires, du niveau de transcription des rétrotransposons chez la vigne. Le chapitre 1 présente les résultats de la recherche de rétrotransposons parmi les séquences EST (Expressed Sequence Tag) de vigne disponibles dans les banques de données. Le chapitre 2 regroupe les résultats de l'évaluation, par RT-PCR semi-quantitative, du niveau de transcription des rétrotransposons dans les organes au cours du développement, et dans les feuilles en réponse à un stress biotique ou abiotique. Et le chapitre 3 présente l'évaluation de la taille des transcrits de quatre familles par Northern blot.

La partie IV présente, sous la forme d'un article, les résultats de l'étude structurale de la famille *Tvvl*, et discute ensuite de l'appartenance à cette famille d'un rétrotransposon identifié dans le génome de la pomme de terre sauvage.

La partie V regroupe les protocoles qui n'ont pas été détaillés dans les chapitres précédents. Finalement, le manuscrit se termine par une discussion générale qui souligne l'intérêt de ce travail et présente les perspectives envisagées.

Partie II : Les rétrotransposons du génomme de la vigne

Chapitre I : Caractérisation de nouvelles familles de rétrotransposons à LTR

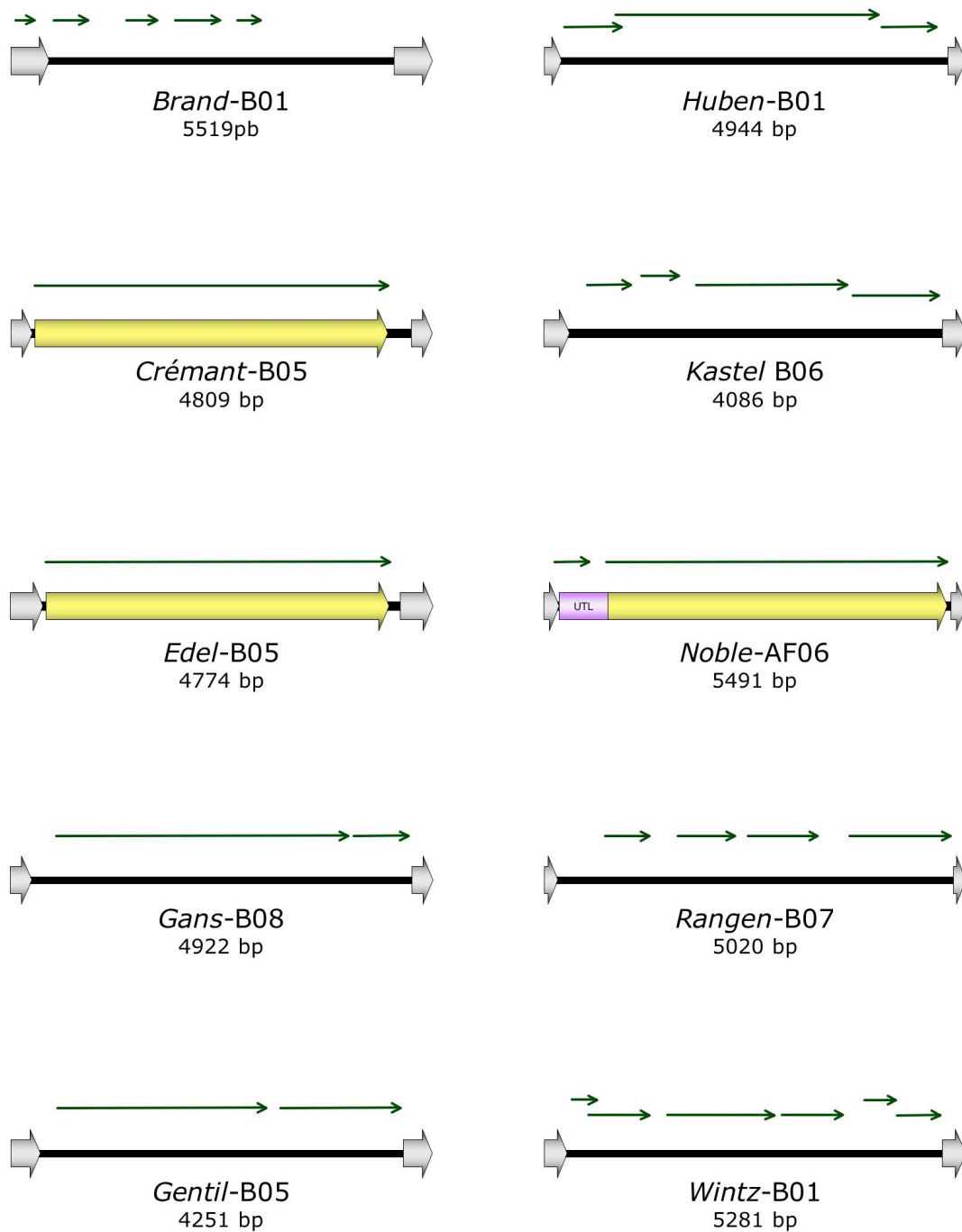
Au début de ma thèse, les informations disponibles sur les éléments transposables du génome de la vigne étaient très limitées puisque seuls trois rétrotransposons étaient décrits. Ainsi, une partie de mon travail a consisté à caractériser de nouvelles familles de rétrotransposons à LTR, pour apporter des informations sur la diversité structurale de ces éléments chez la vigne et étudier leur activité transcriptionnelle.

I.1 - Méthode de Pearce *et al.*

En 2003, Keith Garrison, du Département Viticulture et Oenologie de l'Université de Californie (USA), a apporté au laboratoire 27 séquences correspondant à la région RNaseH-PPT-LTR de plusieurs familles de rétrotransposons de vigne. Ces séquences ont été obtenues en utilisant la stratégie proposée par Pearce et collaborateurs pour caractériser des séquences terminales de rétrotransposons Ty1-*copia*-like. Ils ont en effet décrit une méthode qui permet l'amplification de fragments de rétrotransposons chez plusieurs plantes, entre un domaine conservé de la région RNaseH et un site de restriction présent dans les LTR (Pearce *et al.*, 1999). Pearce et collaborateurs ont ainsi obtenu plusieurs séquences correspondant à la région RNaseH-PPT-LTR de plusieurs familles de rétrotransposons Ty1-*copia*-like du pois, de la fève et de l'épicéa. Cette stratégie a récemment été utilisée pour identifier le rétrotransposon *Tmt1* du génome de la truffe (Riccioni *et al.*, 2008).

I.2 - Distinction de 10 nouvelles familles de rétrotransposons à LTR

Les premières données exploitables du génome de la vigne étaient des séquences de quelques kilobases obtenues à partir du séquençage du génome du Pinot noir. Nous y avons recherché par Blast les éléments correspondant aux 27 fragments que nous avons obtenus précédemment. Rapidement, nous nous sommes aperçus que les copies de rétrotransposons



→ = cadre de lecture ouvert, — = ADN non-codant, → = Séquence codant potentiellement pour une polyprotéine *gag-pol* complète, ⇨ = LTR, UTL = UTL (Untranslated Leader region) de taille variable entre les copies de la famille

Figure 27 : Principaux cadres de lectures ouverts détectés dans les copies canoniques des 10 nouvelles familles de rétrotransposons à LTR du génome de PN40024.

que nous obtenions à partir de ces séquences étaient incomplètes, ou ne présentaient pas les mêmes duplications du site d'insertion en 5' et en 3'. Ce cas de figure peut parfois être le résultat de recombinaisons entre deux éléments de la même famille (Sabot and Schulman, 2007) mais il s'avérait être très fréquent dans ces séquences. Nous en avons donc conclu que les rétrotransposons identifiés dans les séquences du génome du Pinot noir étaient presque systématiquement des chimères résultant d'un problème d'assemblage des contigs, sans doute provoqué par le taux d'hétérozygotie élevé que présente ce génome (environ 75%).

Dès que les premières séquences du génome de PN40024 (homozygote à 93%) ont été disponibles, nous avons recherché les éléments correspondant aux 27 fragments RNaseH-PPT-LTR et nous avons pu identifier 21 rétrotransposons complets, encadrés par des direct repeats identiques. L'analyse phylogénétique de ces éléments nous a ensuite permis de caractériser 10 nouvelles familles de rétrotransposons à LTR, qui se distinguent des trois éléments déjà connus chez la vigne et des rétrotransposons décrits chez les autres plantes. Ces nouvelles familles ont été baptisées à partir des noms de vins et de lieux de production de vin en Alsace : *Brand*, *Crémant*, *Edel*, *Ganz*, *Gentil*, *Huben*, *Kastel*, *Noble*, *Rangen* et *Wintz*.

La taille et le nombre de séquences du génome de PN40024 disponibles augmentant progressivement, nous avons pu caractériser un nombre plus important de copies appartenant à chaque famille. Pour chacune d'entre elles, nous avons notamment recherché une séquence canonique, devant présenter les caractéristiques structurales attendues chez des copies matrices, c'est-à-dire des séquences de l'hôte dupliquées de part et d'autre de l'élément au cours du processus d'insertion ; deux LTRs très similaires ; et un cadre de lecture codant pour une polyprotéine *gag-pol* potentiellement fonctionnelle. Les copies canoniques des 10 nouvelles familles ont des tailles comprises entre 4086 et 5519 pb, et des LTRs de 147 à 502 pb. Des copies potentiellement actives et autonomes, qui présentent un cadre de lecture complet et non-interrompu par un codon stop, ont été détectées pour les familles *Crémant*, *Edel* et *Noble* (Figure 27). Pour les autres familles, aucune copie présentant un cadre de lecture ouvert suffisamment grand pour coder une polyprotéine complète, c'est-à-dire portant tous les domaines fonctionnels attendus chez les rétrotransposons, n'a été détectée.

Ne disposant pas d'outil informatique adapté à l'annotation systématique des rétrotransposons, la caractérisation des copies complètes de chaque famille a été réalisée *manuellement*, mais nous avons finalement pu caractériser plus de 600 copies complètes,

correspondant aux 10 nouveaux rétrotransposons ainsi qu'à *Gret1*, *Tvv1* et *Vine-1*. Nous avons ensuite localisées ces copies sur les différents chromosomes de PN40024.

L'estimation du temps écoulé depuis l'insertion des copies dans le génome, basée sur la divergence des LTRs des copies complètes, montre que chaque famille a probablement connu une dynamique d'amplification particulière, certaines familles s'étant amplifiées récemment dans le génome. Au total, près de 1700 copies, complètes ou partiellement dégénérées, appartenant à ces 13 familles occupent environ 2% du génome de *V. vinifera*.

Les résultats de l'ensemble de cette étude sont présentés sous la forme d'un article en cours de rédaction pour le journal « BMC Genomics ».

Characterisation of ten novel Ty1 *copia*-like retrotransposon families of the grapevine genome

Cédric Moisy¹, Keith Garrison^{2,3}, Carole Meredith², Didier Merdinoglu¹, Frédérique Pelsy¹

¹UMR1131 Santé de la Vigne et Qualité du Vin, INRA, Université Louis Pasteur de Strasbourg, F-68000 Colmar, France

² Department of Viticulture and Enology, University of California, Davis, CA95616,

³ Department of Medicine, Division of Experimental Medicine, University of California, San Francisco, CA 94143 USA

Communicating author: Frédérique Pelsy.

E-mail pelsy@colmar.inra.fr

Tel : + 33-389-224946

Fax : + 33-389-224933

Abstract

Retrotransposons make a significant contribution to the size, organization and genetic diversity of genomes. To characterize retrotransposon families in the grapevine genome, the fourth crop plant genome being sequenced, we have combined two strategies: an RNaseH-LTR analysis of sequences isolated with the Pearce technique and a computer-based sequence similarity search in the whole-genome sequence of PN40024. Supported by a phylogenetic analysis, ten novel families were distinguished. For each of them, a canonic element was identified by looking for the presence of perfect 5 bp-duplication target sites, the highest LTR identity and the most probable deduced ORF.

One to eight copies encoding a single putatively functional *gag-pol* polyprotein were identified for three families, indicating that these families could be still autonomous and active. For the others, no autonomous copies were identified, although some of them displayed perfect LTRs indicating a recent insertion. It could be possible that PN40024, an almost homozygous accession, has lost the corresponding active copies during the self-fertilization successive cycles leading to its homozygosity level. However, these functional copies could still remain in the heterozygous parent and in other grapevine accessions as well. A phylogenetic analysis study based on the sequence alignment of the region located between reverse transcriptase domains I and VII distinguished these 10 families from other plant retroelements. Based on their LTR divergence, insertion dating of full-length copies belonging to 13 grapevine families shows that each family has a particular amplification history, 71% of the copies being inserted within the last 2 million years.

Background

Sequencing of whole genomes shows the predominant amount of DNA composed of transposable elements that shape them. Four plant genomes are now fully sequenced, those of Arabidopsis [1], rice [2], black cottonwood [3] and more recently grapevine [4]. Annotation revealed that amount of repetitive DNA depend on each species, composing ~10% of the Arabidopsis genome (~125 Mbp), ~35% of rice genome (389 Mbp), and ~41.4% of the grapevine genome (487 Mbp). Analysis of repeated DNA sequences carried out in these genomes and in larger cereal genomes as well have shown that insertion of long terminal repeat (LTR) retrotransposons of the two main superfamilies, Ty3 *gypsy*-like and Ty1 *copia*-like are the major components of intergenic regions. Retrotransposons are mobile elements closely related to retroviruses in their structure and life cycle [5]. Active retrotransposons transpose via the reverse transcription of a transcribed, intermediate RNA, thus increasing their copy number [6]. The copy-and-paste cycle in the host genome comprise the synthesis of specific proteins encoded by two major genes, *gag* and *pol*. The Ty1 *copia* [7, 8] and the Ty3 *gypsy* [9] superfamilies differ from each other in the order of domains inside the *pol* gene as well as in their degree of sequence conservation. The *pol* gene encodes from 5' the protease, integrase, reverse transcriptase (RT), and RnaseH domains. These enzymatic functions are required to synthesize an extrachromosomal DNA daughter copy from the intermediate RNA prior to its reinsertion into the genome. The process of reverse transcription being highly error-prone due to the lack of proofreading [10], the daughter copies can differ from the master copy. Moreover, since retrotransposons are believed to be free from selection pressure, copies can diverge independently by mutations after their insertion. Therefore, a hallmark of retroelements is the genetic heterogeneity of individual copies belonging to a same family [11], the RT domain having the slowest relative rate of change among all retroelements proteins while the *gag* domain being more than twice higher [12]. It has been propose that a particular element can be grouped into a family when its amino acid identity is $\geq 90\%$ between reverse transcriptase domains I and VII and into a superfamily when it is $\geq 75\%$ [13].

Because of the mechanism of transposition, the two LTRs of a retrotransposon are theoretically identical at the time of their insertion [14-16]. But after insertion LTRs start evolving independently and diverging in sequence. Thus, the sequence divergence of LTRs in individual elements can be used to estimate when their insertions at the locus took place [17].

Different strategies have been used for characterize retrotransposons. Full-length or parts of Ty1 *copia*-like retrotransposon were trapped into coding sequence, such as *Tnt1* in the tobacco nitrate reductase gene [18] or *Vine-1* in a grape *Adhr* gene [19]. Other can determine phenotypic character such as the color of grape skins associated with the presence of the *Gret1* retrotransposon in the promoter region of *VvMybA1* [20]. Retroelement sequences have also emerged from computer-based sequence similarity searches in genomic databases [21]. To rapidly deliver Ty1 *copia*-like retrotransposon LTR sequences of any higher plant, Pearce *et al* have used a PCR-based method to isolate novel LTR sequences from pea, broad bean and Norway spruce [22]. Following this procedure, Riccioni *et al* (2008) have isolated Tmt1, the first LTR-retrotransposon from ectomycorrhizal fungi of the genus *Tuber* that shows relatedness to Ty3 *gypsy* retrotransposons.

Because retrotransposons make a significant contribution to the size, organization and genetic diversity of genomes, grapevine retrotransposons are important to be characterized. Three retrotransposons were described before the grape whole-genome sequencing: a unique copy of *Vine-1*, a Ty1 *copia*-like element [19] and of *Gret1*, a Ty3 *gypsy*-like element [20] and the *Tvv1* family [23]. The grape sequence recently been obtained from the highly homozygous genotype PN40024 is a genomic resource that makes possible the characterization *in silico* of the grape retrotransposons. Annotation of grapevine transposable elements revealed a large prevalence of retrotransposons over transposons, most of which (47%) correspond to Ty1-*copia* like retrotransposons and their uneven distribution of over the grape chromosomes, mainly restricted to low gene density regions and introns [4]. In this paper, we described the characterization of novel families of Ty1 *copia*-like of the grapevine genome associating an RNaseH-LTR analysis of sequences isolated with the Pearce technique and a computer-based sequence similarity search. Supported by phylogenic analysis novel families were distinguished and for each one a canonic element was identified. Then, their phylogenic relationship with reference plant retrotransposons was studied. Based on their LTR divergence, insertion dating of full-length copies belonging to 13 grapevine families was estimated in order to study the dynamic of these families.

Methods

Plant material and DNA extraction

3'-end retroelement sequences were isolated from pinot noir S15. Total genomic DNA extracts were performed using the Qiagen Plant extraction kit (Qiagen, Valencia, CA) according to manufacturer's protocols.

Sequence isolation

Isolation of 3'-end retroelement sequences was performed according to Pearce et al. (1999) with modifications. The first amplification was performed using an adapter primer containing a single selected nucleotide used in pair with a primer designed in RNaseH motif 1. It was followed by a nested amplification performed using the same adapter primer in pair with a primer designed in RNaseH motif 2. PCR products were cleaned up using the QIAquick PCR purification kit (Qiagen, Valencia, CA) before being cloned using the Promega pGEM T-Easy cloning kit (Promega, Madison, WI) with Invitrogen DH5 E. coli cells (Invitrogen, Carlsbad, CA) according to manufacturer's protocols.

Sequence analysis

Computer-assisted analysis of the 3'-end retroelement sequences fragments was performed in DNASTAR SeqManII. The nucleotide data of the 24 reported elements have been deposited in GenBank database under accession numbers XXX. Full-length corresponding elements were searched in the grapevine genome using Genoscope BLAST server (http://www.cns.fr/cgi-bin/blast_server/projet_ML/blast.pl) looking for LTRs separated by 4-5 Kb and for target duplication sites which mark retrotransposon insertion. They were then individually examined for position of their LTRs, UTR, ORFs in DNAsis 2.1 (Hitachi Software Engineering Co, Ltd) and Vector NTI 7 (InforMax Inc). MEGA v3.1 [24] was used to achieve multiple alignment of the complete nucleic sequence by Clustal W and to draw Neighbor-Joining tree using Jukes and Cantor model [25]. Amino acid sequences between reverse transcriptase motifs I to VII were deduced from the nucleic acid sequences and when needed, they were constructed from the hypothetical translation of the three ORFs of the internal sequence. They were compared two by two using the EMBOSS Pairwise Alignment Algorithms (<http://www.ebi.ac.uk/emboss/align/>). The 10 families were named after wine produced or vineyards cultivated in Alsace country.

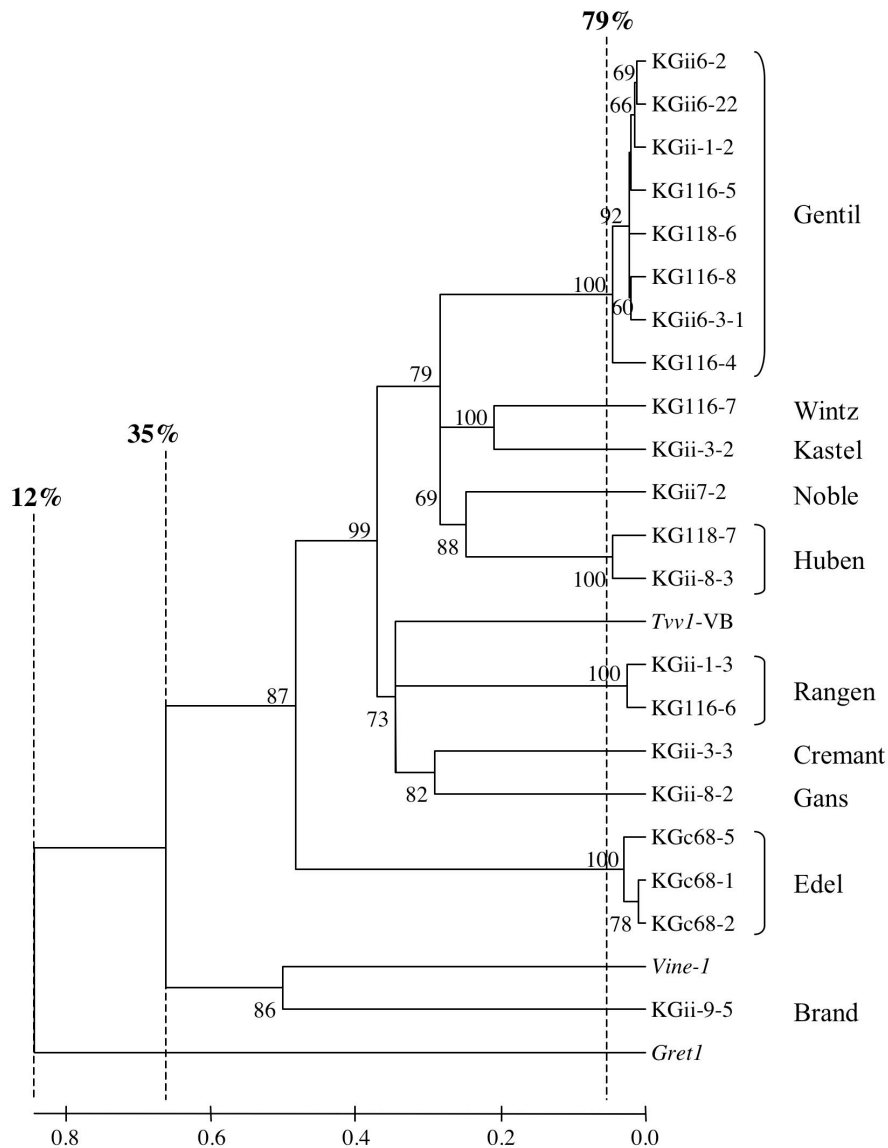


Figure 1: Classification of 21 mined full-length retrotransposons in 10 families. The Neighbor-Joining tree is based on the alignment of the complete nucleic sequence and distance scales correspond to a percentage of nucleotide substitution under the Jukes and Cantor model [25]. The lowest identity percentages within or between families were given. Identity levels were estimated comparing two by two corresponding amino acid sequences between reverse transcriptase domains I and VII. *Tvv1*, *Vine-1* and *Gret1* were added as reference. Bootstrap values greater than 60% are given on the tree.

Data mining

For each family, the whole sequence of a reference copy was chosen to perform a BLAT search against the PN40024 genome (<http://www.genoscope.cns.fr/blat-server/cgi-bin/vitis/webBlat>). BLAT on DNA is designed to quickly find sequences displaying stretches 40 bases or more in length 95% and greater similar to query. It also finds perfect shorter sequence matches. All matches that had a length higher than 80% and a similarity level higher than 75% of the reference were considered to identify canonical copies. Using *gag-pol* polyprotein sequence available from public databases, we selected aminoacid sequences between reverse transcriptase motifs I to VII. A phylogeny reconstruction was achieved using MEGA v3.1 to perform the multiple sequence alignments and build a Neighbor-Joining tree based on the Poisson model.

Genomic paleontology

The insertion of full-length copies were dated based on the divergence of the 5' and 3' LTR sequence of each copy, as proposed by San Miguel et al. [17]. LTR sequences were aligned MEGA v3.1 and the observed divergence was corrected following the Jukes and Cantor molecular evolution model, as proposed by Vitte et al. [26]. This corrected divergence was then translated to estimate the insertion date of the elements at their present location, using the average base substitution rate derived from the grass *adh1-adh2* region of 6.5×10^{-9} substitution per site per year [27], given the assumption of a uniform mutation rate in grapevine.

Results

Characterization of ten Ty1 copia-like retrotransposon families of the grapevine genome

Applying the technique of Pearce *et al.* (1999) to the grapevine genome resulted in the isolation of 27 fragments displaying different sequences. Among them 24 contained an identifiable portion of the 3'-end of Ty1 *copia*-like retrotransposons including a putative RNaseH domain, polypurine tract, and varying amounts of the 3'LTR sequence. Eighteen of them were unique and six almost identical (further characterized as *Edel* sequences). None of these fragments were similar to one of the three grapevine retrotransposon families previously characterized.

A data mining procedure using the Blaster algorithm with the 18 unique and three among the almost identical sequences as queries yielded 21 different full-length retrotransposons harboring an internal domain flanked by LTRs, twelve of them displaying perfect target

Family name	Brand	Cremant	Edel	Gans	Gentil	Huben	Kastel	Noble	Rangen	Wintz	Gretl	TvvI	Vine-I	Total
Reference copy	KGii9-5	KGii-3-3	KGc68-1	KGii-8-2	KGii6-2	KGii8-7	KGii-3-2	KGii7-2	KGii-1-3	KGii16-7	ABii11100	TvvI-IB	AFii16598	
Reference size (pb)	5519	4807	4783	4217	4265	4941	4072	5502	5101	5278	10422	5222	2392	
Copy number	92	110	187	94	212	91	91	185	105	126	173	103	140	1,709
Identity (%)	70.9	54.0	93.2	80.0	62.9	60.0	88.5	86.3	60.8	62.1	74.5	62.8	91.6	
	87.75	91.05	96.6	97.1	95.7	90.7	93.7	97.3	94.4	80.35	88.3	97.4	96.9	
	85.74	89.69	96.35	96.08	95.35	90.93	93.60	97.25	91.78	81.68	88.95	94.57	96.59	
Size (pb)	100	100	99.9	100	100	100	100	100	100	100	100	100	99	
	22	17	74	19	16	17	16	26	16	20	21	26	20	
	633	45	4407	253	3088	2922	3623	5122	2180	616	6290	245	292	
	1585	589	4001	1520	2834	2589	3180	4229	2590	903	5944	2117	1127	
	9177	6921	18308	4960	4852	5824	7410	7040	7113	5396	10423	6963	2573	
Size (% reference)	0.4	0.4	1.5	0.5	0.4	0.3	0.3	0.5	0.3	0.3	0.2	0.5	0.8	
	11.5	0.9	92.1	6.0	72.4	59.1	88.9	93.1	42.7	11.7	60.4	4.7	12.2	
	28.7	12.3	83.6	36.1	66.5	52.4	78.1	76.9	50.8	17.1	57.0	40.5	47.1	
	166.3	144.0	382.8	117.6	113.8	117.9	182.0	128.0	139.4	102.2	100.0	133.3	107.6	
Location PN40024	chr 1	5	7	5	15	7	3	11	3	9	4	3	4	73
	chr 2	4	4	4	7	3	4	9	3	11	7	1	4	57
	chr 3	1	4	4	8	3	3	8	3	1	2	2	33	
	chr 4	8	1	9	8	1	6	10	1	6	5	3	2	60
	chr 5	1	2	8	10	7	4	7	7	5	10	8	7	86
	chr 6	6	5	5	5	1	2	5	7	2	16	9	3	69
	chr 7	4	4	2	9	4	4	3	2	1	7	11	9	50
	chr 8	1	2	5	5	5	3	4	9	2	6	13	4	54
	chr 9	3	10	10	12	4	6	5	8	1	10	6	6	68
	chr 10	4	6	8	3	5	1	6	8	8	3	4	1	63
	chr 11	4	8	2	4	3	5	7	1	9	3	2	4	52
	chr 12	12	7	9	10	3	2	6	7	10	5	4	7	91
	chr 13	8	2	9	2	2	2	8	4	4	4	2	13	61
	chr 14	10	8	6	3	9	5	10	9	5	6	5	5	95
	chr 15	1	9	8	9	9	4	4	4	2	5	2	5	51
	chr 16	3	11	7	3	1	5	9	3	3	1	2	6	59
	chr 17	4	5	7	3	3	4	4	2	5	15	3	5	63
	chr 18	5	11	10	8	4	8	18	4	7	13	8	10	122
	chr 19	6	7	7	2	1	1	3	4	9	8	2	8	64
Unknown	17	17	60	23	51	34	21	48	29	26	43	23	46	438
Inserted DNA (Kb)	507.75	528.77	894.42	396.40	904.18	449.63	370.55	1017.87	535.61	665.03	1803.1	537.87	334.88	8,946.06
Coverage (Kb)	145.86	64.80	748.11	142.92	600.89	235.59	289.34	782.34	271.94	113.75	1028.36	218.08	157.83	4,799.81
% PN40024 genome	0.03	0.01	0.15	0.3	0.12	0.05	0.06	0.16	0.06	0.02	0.21	0.04	0.03	1.24

Table 1: Global analysis of the 10 grapevine retrotransposon families
Copy number refers to elements which similarity is upper than 75% on more than 80% in length with the reference sequence. Min : minimum ; med : medium ; Max : maximum ; Chr : chromosome

duplication sites. Then, a phylogenetic analysis was conducted using these 21 elements in addition with *Tvv1*, *Vine-1* and *Gret1*, three retrotransposons previously characterized in the grapevine genome. The Neighbor-Joining tree based on the alignment of their complete nucleic sequence clustered the 21 elements in 10 distinct families different from *Tvv1*, *Vine-1* and *Gret1* (figure 1). The largest family *Gentil* grouped eight elements, the *Edel* family three elements and two families, *Huben* and *Rangen*, two elements. Finally, six families, *Brand*, *Cremant*, *Gans*, *Kastel*, *Noble* and *Wintz* were represented by a single element. Corresponding amino acid sequence domain research confirmed the order of domains inside the *pol* gene specific of Ty1 *copia*-like elements. To confirm the clustering in 10 distinct families, the amino acid sequence located between reverse transcriptase domains I and VII of the 21 elements were compared. Identity levels evaluated as suggested by [13] was $\geq 79\%$ within elements classified in the same family, ranged 67-35% between elements classified in distinct Ty1 *copia* families and 21-12% between *Gret1*, a gypsy element, and the Ty1 *copia* families.

Representation of the 10 families in the grapevine genome

The BLAT program was used to extract the paralogous copies of the ten families from the grapevine genome along with *Tvv1*-VB, *Vine-1* and *Gret1*. A total of 1709 copies were identified, the copy number per family ranging from 212 to 91. Families comprising the higher copy number were *Gentil* (212), *Edel* (187), *Noble* (185), and *Gret1* (173) while *Huben* and *Kastel* (91) had the lowest copy number (Table 1). For all families except *Gret1* copies longer than the reference size were found. In particular one *Edel* element displaying a long insertion contributes to extend his size over 382% of the reference copy. More generally, all families have a median length below the reference query indicating that most of the copies should be truncated, however, with a high sequence identity (median range 0.80-0.97). Using *Vine-1* as query, most of the occurrences correspond to smaller sized copies. The larger copy being 107.6% of the reference *Vine-1* could not encode the RT-RnaseH domain lacking in the reference *Vine-1* element [19].

Considering that all mined copies could have had the size the corresponding reference sequence upon insertion, they would have represented a total DNA quantity of 8.9Mb added to the genome by their specific amplification. The *Gret1* family, which reference copy sizes 10,422 bp, and *Noble* among Ty1 *copia* families might have contributed to the larger part of DNA content. Summing all element current lengths, a residual amount of 4.8 Mb represents 1.24% of the 40024 genome suggesting that 47% had been removed. Copies classified in the 13 families are scattered around chromosomes as currently assembled with a main density of

Location PN40024	Total copy number	Chromosome size (Mb)	rank/size	density (copy/ Mb)
chr 1	73	15.63	10	4,67
chr 2	57	17.6	8	3,24
chr 3	33	10.19	16	3,24
chr 4	60	19.29	6	3,11
chr 5	86	23.43	2	3,67
chr 6	69	24.15	1	2,86
chr 7	50	15.23	11	3,28
chr 8	54	21.56	3	2,50
chr 9	68	16.53	9	4,11
chr 10	63	9.647	17	6,53
chr 11	52	13.94	14	3,73
chr 12	91	18.54	7	4,91
chr 13	61	15.19	12	4,02
chr 14	95	19.48	5	4,88
chr 15	51	7.694	19	6,63
chr 16	59	8.159	18	7,23
chr 17	63	13.06	15	4,82
chr 18	122	19.69	4	6,20
chr 19	64	14.07	13	4,55
Unknown	438			

Table 2: Copy density of the on the 19 current chromosomes

	Element size (bp)	5'LTR (bp)	Internal domain (bp)	3' LTR (bp)	LTR identity %	Direct flanking repeat	LTR end dinucleotides	UTL size (bp)	gag-pol (aa)	Location	NCBI accession n°
Brand-B01	5519	500	4517	502	94.4	AATAT	TGTT/AACA			chrUn_random:34708921..34719959 scaffold_163	CU459380
Cremant-B05	4809	245	4320	244	99.2	ATATA	TAAC/GTCA	35	1346	chr11_random:1632133..1636941 scaffold_118	CU459335
Edel-B05	4774	373	4028	373	100	GGAGC	TGTT/TACA	39	1298	chr8:16967018..16976565 scaffold_23	CU459240
Gans-B08	4922	247	4428	247	100	GTAGT	TGTT/TACA			chr12:17929297..17939140 scaffold_78	CU459295
Gentil-B05	4251	297	3657	297	99.7	GTTGT	TGTT/AAGA			chr9:11377491..11385993 scaffold_33	CU459250
Huben-B01	4944	206	4532	206	100	ATATA	TGTG/TCA			chr14:16202771..16212658 scaffold_21	CU459238
Kastel-B06	4086	249	3606	231	89.2	GTAA	TGTT/GTCA			chr5:17548931..17553016 scaffold_67	CU459284
Noble-AF06	5491	201	5089	201	100	CAAC	TGTT/TCA	621	1486	chr5:17660817..17666309 scaffold_21	CU459238
Rangen-B07	5020	168	4705	147	91.9	AATGC	TGTT/AACA			chr19:12572975..12577994 scaffold_35	CU459252
Wintz-B01	5281	291	4999	291	99.3	AAAC	TGTG/GTCA			chr2:15397291..15402571 scaffold_140	CU459357

Table 3: Sequence characteristics and location of the ten canonical elements identified in the grapevine genome sequence

4.43 copies per Mb (Table 2). However, this distribution is not uniform. Displaying 54 copies, chromosome 8 (21.56 Mb) has the lower density (2.50) while chromosome 16 (8.159 Mb) has a density almost third higher (7.23). Most of families seem to be rather uniformly distributed on the chromosomes except *Gans* absent of 6 chromosomes but present in 17 occurrences on chromosome 5.

Identification of the canonical copies

For each family, a canonical copy was identified among elements which length was at least 80% of the reference size with an identity level higher than 75%. Perfect 5 bp-duplication target sites, high LTR identity and an ORF potentially encoding a full-length *gag-pol* polyprotein should characterize these copies. The 10 canonical elements selected varied in size from 4,086 to 5,519 bp, with LTRs ranging from 147 to 502 bp (table 3). Eight of the family LTRs begin and end with the dinucleotide inverted repeat TG...CA as observed for many LTR-retrotransposons and retroviruses (Dej et al. 1998). More precisely, seven start with the four-nucleotide stretch TGTT, among them two, *Brand* and *Rangen*, end with the corresponding inverted stretch AACAA. *Huben* and *Wintz* LTRs begin with the four-nucleotide stretch TGTG and end with TTCA and GTCA, respectively. Finally, *Cremant* LTRs begin with the four-nucleotide stretch TAAC and end with GTCA.

One *Cremant*, 3 *Edel* and 8 *Noble* elements showed a single putative ORF uninterrupted by a stop-codon corresponding to polyproteins sizing 1,346, 1,298 to 1,486 amino-acids long respectively, indicating that they could be autonomous. By comparison with *Ty1 copia*-like plant retrotransposons previously characterized, the general organization of the putative ORFs of *Cremant*, *Edel* and *Noble* elements is very similar both in structure and in length, compared to the 1,296, 1,382, 1,328 and 1,447 amino acid-long ORF of the barley *BARE-1* (acc. n° Z17327), grapevine *Tvvl-VB* (acc. n° EU304807), tobacco *Tnt1* (acc. n° X13777), and *Arabidopsis thaliana AtRE1* (acc. n° AB021265) retrotransposons, respectively. The putative ORFs of the seven remaining elements were prematurely interrupted by a stop codon or a frameshift in the coding regions. Although *Gans*, *Gentil*, *Huben* and *Wintz* canonic copies lack a putative full-length polyprotein, the homology between LTRs (100%, 99.7%, 100% and 99.3%, respectively) indicates that they moved recently at their insertion site (table 3). The 8 *Noble* copies putatively active displayed an ORF 4,458 bp-long but differed by their UTL region sized from 578 to 632 bp-long. Moreover, taking in consideration the 40 full-length *Noble* copies attested by perfect host 5bp-target sites, their UTL region varied from 269 bp to 695 bp (average 610 bp), all started by the TGGTATCAGAGCC which correspond

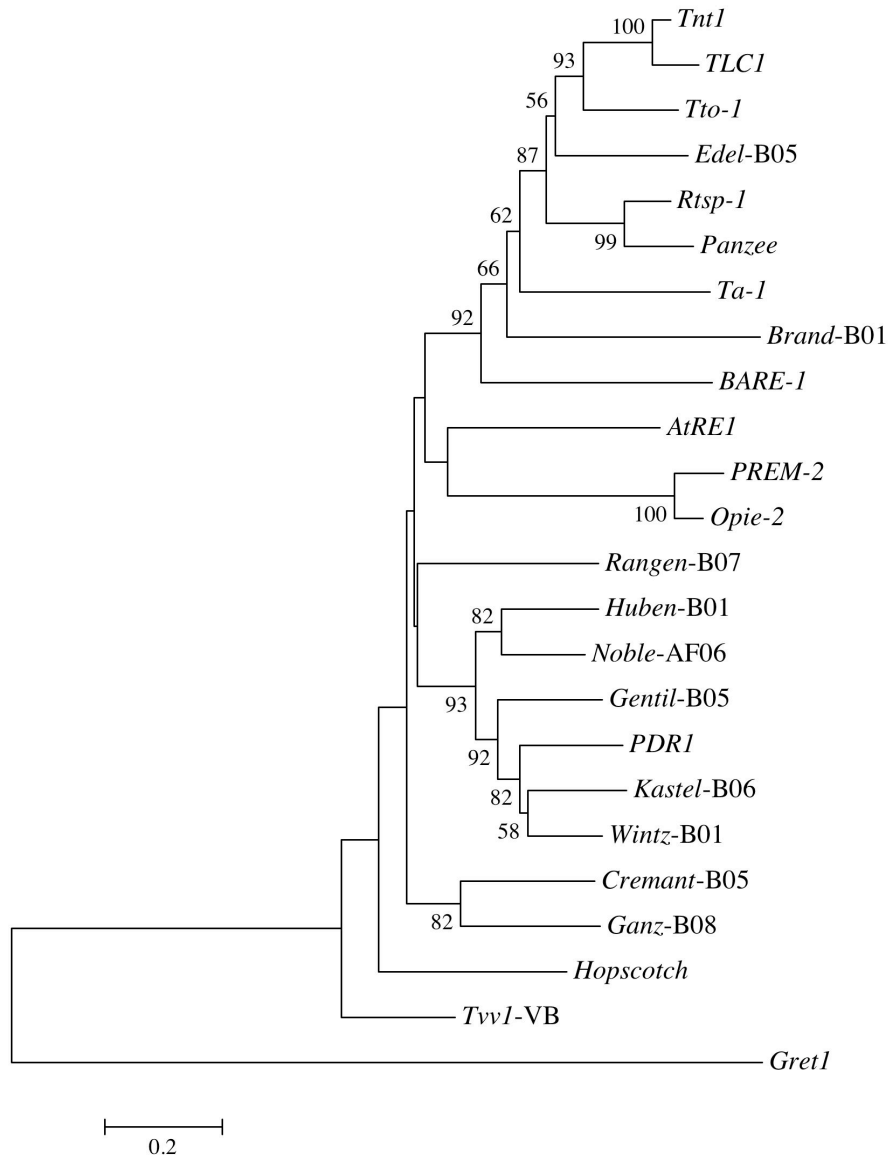


Figure 2: Phylogeny analysis of the 10 novel families.

The Neighbor-Joining tree is based on the multiple sequence alignment of amino acid sequences between reverse transcriptase domains I and VII using the Poisson distance model. For novel families, amino acid sequences were deduced from canonic copies and for other plant retrotransposons from released nucleic sequences: *Nicotiana tabacum Tnt1* (X13777); *Solanum chilense TLC1* (AF279585); *N. tabacum Tto1* (D83003); *Ipomoea batatas Rtsp-1* (AB162659); *C. cajans Panzee* (AJ000893); *A. thaliana Ta-1* (X13291); *Hordeum vulgare BARE-1* (Z17327), *A. thaliana AtRE1* (AB021265); *Zea mays PREM-2* (U41000); *Z. mays Opie-2* (AF090446); *Pisum sativum PDR1* (X66399); *Z. mays Hopscotch* (U12626); *V. vinifera Tvv1-VB* (EU304807); *V. vinifera Gret1* (AB111100). Bootstrap values greater than 50% are given on the tree.

to the PBS. A detailed analysis of 12 UTL *Noble* indicated that variations in size have essentially resulted from large deletions/insertions within the internal region as already described for the UTL region of *Tvv1* [28]. To elucidate the phylogenetic relationships among the ten novel families, *Tvv1*, *Gret1*, and 14 plant *Ty1 copia*-like, the amino acid sequences between motifs I to VII of the RT were aligned [29]. *Vine 1* that shows no RT domain could not be included in this study. Any of ten novel families could be clustered with a previously identified families (Figure 2).

Genomic paleontology

Dating of insertion of 163 copies considered as full-length and belonging to the 13 families (the 10 previously described along with *Gret1*, *Vine-1*, and *Tvv1*) were estimated by aligning the 5' and 3' LTR sequence according to Vitte et al.[26]. The histograms (figure 3) revealed that copy age ranges from 11 to 0,2 million years (My). One hundred and fifteen copies (70%) were inserted within the last 2 My among them 45 copies belonging to seven families (27%) that displayed perfect LTRs have moved recently. Two types of histogram were obtained reflecting distinct evolutionary histories and activities of particular families: some families comprise copies of inserted at different date such as *Edel* (5-1 My) or *Gentil* (7-1 My) whereas other families show many copies inserted at the same date such as *Tvv1* (7, 3 My and recent) or *Noble* (0.5 My and recent). For these last families, copies inserted at the same date could reflect different bursts of amplification.

Discussion

Distinction of the novel families

Combining an RNaseH-LTR analysis of sequences isolated with the Pearce technique and a computer-based sequence similarity search in the whole-genome grapevine sequence, we have identified ten novel *Ty1 copia*-like retrotransposon families, none of them belonging to a grapevine retrotransposon family previously described. This result proves the complementarities of both techniques to isolate retrotransposon 3'-end sequences first and to identify the corresponding family afterwards. However, among the 24 sequences containing an identifiable portion of the 3'-end of retrotransposons, 18 were unique and six almost identical, further shown to correspond to *Edel* elements.

The clustering of the 21 different full-length retrotransposons identified by Blast into distinct families was based both on the alignment of their complete nucleic sequence and of the

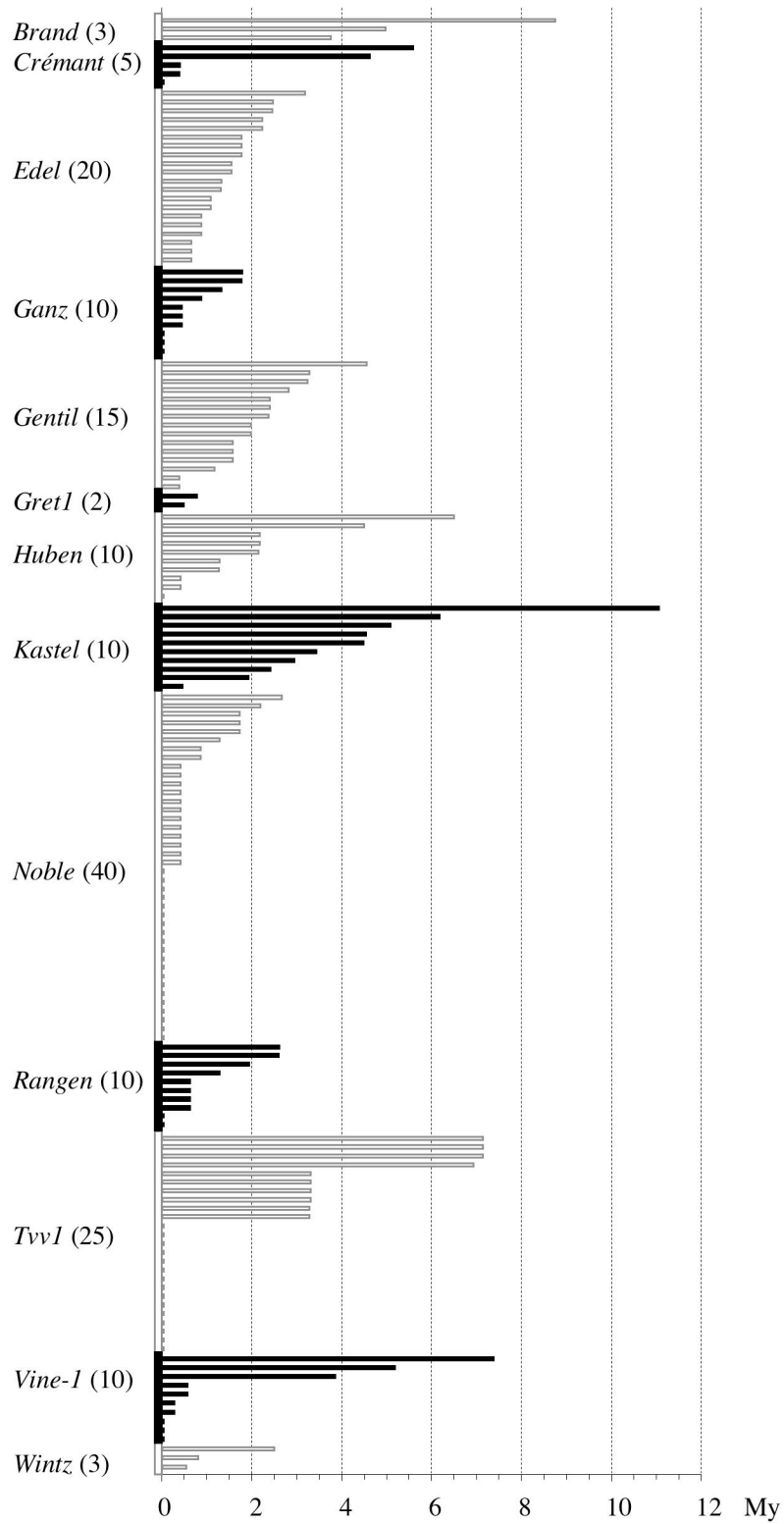


Figure 3: Distribution of the insertion dates for copies belonging to 13 grapevine LTR retrotransposon families.

LTR divergence was used to estimate date of insertion of full-length copies flanked by perfect direct repeats

corresponding amino acid sequences between reverse transcriptase motifs I to VII as suggested by [13]. Ten families were identified comprising from eight to a unique element. Identity level comparing the amino acid sequences is upper than 79% between elements clustered in the same family while it decreases from 67 to 35% between elements classified in distinct Ty1 *copia* families and drops from 21 to 12% between any Ty1 *copia* families and *Gret1*, a unique grapevine *gypsy* element. However considering the same amino acid region, the identity values we obtained are lower than those proposed by [13] that grouped a particular element into a family when its amino acid identity is upper than 90% and into a superfamily when it is upper than 75%. Thus, in our study, identity level values do not deliver a clear-cut information sufficient to classify unknown elements.

Based on the order of domains inside the *pol*, the ten families we identified were all related to of Ty1 *copia* retrotransposons subfamily. Using the same procedure, Tmt1, a *Tuber* LTR-retrotransposon was isolated. Its relatedness to Ty3 *gypsy* retrotransposons superfamily was further established [30]. Moreover, the number of elements classified in a same family reveals an over-representation of elements belonging to the two families *Gentil* and *Edel*. On the contrary, sequences of elements belonging to six families were unique. Finally seven families start with the four-nucleotide stretch TGTT. Thus, the Pearce technique could facilitate the amplification of sequences belonging to particular families, mainly Ty1 *copia* related, to the detriment of others.

Genomic fraction of the genome occupied by 13 grapevine retrotransposon families

The BLAT program used to extract a total of 1709 copies paralogous to the 13 families studied only find sequences of 95% and greater similarity on 40 bases length or more. Because of these stringent conditions parameters, the copy number of these families could possibly be higher. Altogether, the 12 Ty1 *copia*-like families represent 1.03% of the genome compared to the 5.16% identified by BlastX annotation or 8.35% identified by manual annotation of Ty1 *copia* elements in the PN40024 grapevine genome (supplementary table S7). The 1536 occurrences we find corresponding to residual DNA amount of 3,771.48 Kb can be compared to the 17,293 (8.88%) and 56,890 (2.69%) occurrences and 24,640.8 kb (15.30%) and 39,848.3 Kb (9.46%) coverage identified by each annotation, respectively. *Gret1*, the only Ty3 *gypsy* retrotransposon considered in this study covers the larger fraction of the genome, probably due to its length. The 173 identified occurrences cover 1028.36 Kb (0.21%) of the PN40024 genome that can be compared to 17,659.6 Kb and 15,339.8 Kb (5.82% and 6.70%) corresponding to the total fraction of Ty3 *gypsy* identified by both

annotations. However, these 173 occurrences only correspond to 1.79% to 1.22 % of the total Ty3 *gypsy* occurrences depending on the annotation method. It is clear that a large number of grapevine retrotransposons remained to be identified.

Genomic distribution of copies belonging to the ten families is not uniform, chromosome16 having the higher density and chromosome 8 the lower, however, this result has to be cautiously considered since a large number of copies belongs to scaffolds that remain to be attributed to a particular chromosome. This density seems to be independent from the richness in resistance gene analogs (RGA) markers since these two chromosomes are very poor in mapped RGA loci [31]. This result is in agreement with the uneven distribution of transposable elements over the grape chromosomes that is mainly restricted to low gene density regions [4].

Specificity of the different families

A main feature of the novel families, except *Brand*, is their LTR size ranging from 147 to 373 bp, a size shorter than most of the plant retrotransposon previously characterized such as *Tal*, *Tnt1*, and *BARE1* which LTRs length 514, 610 and 1829 bp, respectively. This common feature could then characterize grapevine Ty1 *copia* retrotransposons. Seven families begin with TGTT, a four-nucleotide stretch that also starts many LTR-retrotransposons such as *Tvv1*, *Vine-1*, *Tst1*, *Tto1*, *PDR1*, *Tal* or *copia*. However, neither *Tvv1* nor *Vine-1* sequences were sorted by the Pearce technique in spite of their LTRs sizing 149 and 287 bp, respectively, and starting with the four-nucleotide stretch TGTT. Further analysis finally shows that no full-length copy of *Vine 1*, first described truncated in its 3' ORF sequence [19] were identified in the grapevine genome (table 1).

For each family, a canonical copy was identified by looking for the presence of perfect 5 bp-duplication target sites, the highest LTR identity and the most probable deduced ORF.

Only copies flanked by perfect 5 bp-duplication target sites that result from repair of the integration event were considered to eliminate chimerical copies that could have resulted from errors introduced during genome assemblage. No apparent conservation between 5 bp-direct repeats flanking the 10 canonical elements were observed, only predominance for A-rich sites.

One to eight copies encoding a single putatively functional *gag-pol* polyprotein were identified for three families, indicating that these families could be still autonomous and active. For the others, no autonomous copies were identified, although some of them displayed perfect LTRs indicating a recent insertion. It could be possible that PN40024, an

almost homozygous accession, has lost the corresponding active copies during the self-fertilization successive cycles leading to its homozygosity level. However, these functional copies could still remain in the heterozygous parent and in other grapevine accessions as well. In the *Noble* family, copies differ by their UTL region sized from 269 bp to 695 bp. A sequence analysis determines that this size variation is essentially caused by large deletions/insertions within the internal region, in the same way as for the *Tvv1* elements (Pelsy 2007)

Dating of insertion of the different families

Dating of insertion of full-length copies belonging to 13 families was estimated to provide insights into the evolutionary histories and activities of different families. However, the stringent parameters that we used to mine full-length copies has probably eliminated the most degenerated copies, thus the oldest. The study reveals that each family has a particular amplification history, 71% of the copies being inserted within the last 2 My, possibly thanks to burst of transposition. Moreover, we found very recent copies for seven families among of them we lack to find a master copy with a potentially functional *gag-pol* polyprotein for four of them. It could be possible that PN40024, an almost homozygous accession, could have lost active copies during the self-fertilization successive cycles leading to homozygosity, but that active copies were still be present in the heterozygous parent and also in other grapevine accessions. And other hypothesis could be the trans-activation of inactive copies using active retrotransposon functions. This last hypothesis is proposed to explain recent insertion of *Vine-1* as no full-length *Vine-1* copies were detected by PCR amplification in various varieties (personal data). Finally, chronology of the insertions of the different families, the older *Kastel* copy being ~11 million years old is in agreement with paleo-botanical data that record the first Vitaceae fossils from the Miocene (23.0 to 5.3 millions years) [32].

Acknowledgements

The authors thank Lucie Bevilacqua-Schehrer for valuable technical assistance. Cedric Moisy is a doctoral fellow supported by INRA and the Conseil Régional d'Alsace and Keith Garrison was a doctoral fellow supported by (?). All experiments described in this study comply with the current laws of France.

References

1. Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. *Nature*, 2000: 408: p. 796-815.
2. Sequencing Project International Rice Genome: **The map-based sequence of the rice genome**. *Nature*, 2005: 436(7052): p. 793-800.
3. Tuskan G, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao R, Bhalerao R, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G, Cooper D, Coutinho P, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson D, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai C, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y and Rokhsar D: **The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science*, 2006: 313(5793): p. 1596-1604.
4. The French–Italian Public Consortium for Grapevine Genome Characterization: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla**. *Nature*, 2007: 449: p. 463-467.
5. Boeke J and Corces V: **Transcription and reverse transcription of retrotransposons**. *Annual Review of Microbiology*, 1989: 43: p. 403-434.
6. Feschotte C, Jiang N, and Wessler SR: **Plant transposable elements: Where genetics meets genomics**. *Nature Reviews Genetics*, 2002: 3(5): p. 329-341.
7. Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, and Kumar A: **Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants**. *Nucleic Acids Research*, 1992: 20(14): p. 3639-3644.

8. Voytas DF, Cumming MP, Konieczny A, Ausubel FM, and Rodermel SR: ***Copia-like retrotransposons are ubiquitous among plants.*** *Proceedings of the National Academy of Sciences of the USA*, 1992: 89: p. 7124-7128.
9. Suoniemi A, Tanskanen J, and Schulman AH: ***Gypsy-like retrotransposons are widespread in the plant kingdom.*** *Plant Journal*, 1998: 13(5): p. 699-705.
10. Gabriel A, Willems M, Mules EH, and Boeke JD: ***Replication infidelity during a single cycle of TyI retrotransposition.*** *Proceedings of the National Academy of Sciences of the USA*, 1996: 93(15): p. 7767-7771.
11. Casacuberta JM, Vernhettes S, and Grandbastien M-A: ***Sequence variability within the tobacco retrotransposon Tnt1 population.*** *EMBO Journal*, 1995: 14(11): p. 2670-2678.
12. McClure MA, Johnson MS, Feng DF, and Doolittle RF: ***Sequence Comparisons of Retroviral Proteins: Relative Rates of Change and General Phylogeny.*** *Proceedings of the National Academy of Sciences*, 1988: 85(8): p. 2469-2473.
13. Bowen NJ and McDonald JF: ***Genomic Analysis of Caenorhabditis elegans Reveals Ancient Families of Retroviral-like Elements.*** *Genome Res.*, 1999: 9(10): p. 924-935.
14. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, and Bennetzen JL: ***The paleontology of intergene retrotransposons of maize*** *Nature Genetics*, 1998: 20(1): p. 43 - 45.
15. Jordan IK and McDonald JF: ***Evidence for the role of recombination in the regulatory evolution of Saccharomyces cerevisiae Ty elements.*** *Journal of Molecular Evolution*, 1998: 47(1): p. 14-20.
16. Jordan IK and McDonald JF: ***Tempo and mode of Ty element evolution in Saccharomyces cerevisiae.*** *Genetics*, 1999: 151: p. 1341-1351.
17. San Miguel P, Gaut BS, Tikhonov A, Nakajima Y, and Bennetzen JL: ***The paleontology of intergene retrotransposons of maize.*** *Nature Genetics*, 1998: 20(1): p. 43-45.
18. Grandbastien M-A, Spielmann A, and Caboche M: ***Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics.*** *Nature*, 1989: 337(6205): p. 376-380.
19. Verriès C, Bès C, This P, and Tesnière C: ***Cloning and characterization of Vine-1, a LTR- retrotransposon-like element in Vitis vinifera L., and other Vitis species.*** *Genome*, 2000: 43(2): p. 366-376.

20. Kobayashi S, Goto-Yamamoto N, and Hirochika H: **Retrotransposon-induced mutations in grape skin color.** *Science*, 2004: 304: p. 982.
21. Le QH, Wright S, YU Z, and Bureau T: **Transposon diversity in *Arabidopsis thaliana*.** *Proceedings of the National Academy of Sciences of the USA*, 2000: 97(13): p. 7376-7381.
22. Pearce S, Stuart-Rogers C, Knox M, Kumar A, Ellis N, and Flawell A: **Rapid isolation of plant TY1-copia group retrotransposon LTR sequences for molecular marker studies.** *The Plant Journal*, 1999: 19(6): p. 711-717.
23. Pelsy F and Merdinoglu D: **Complete sequence of *Tvv1*, a family of *Ty1 copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking.** *Theoretical and Applied Genetics*, 2002: 105: p. 614-621.
24. Kumar S, Tamura K, and Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Briefings in Bioinformatics* 2004: 5: p. 150-163.
25. Jukes TH and Cantor CR, *Evolution of protein molecules*, in *Mammalian Protein Metabolism*, H.N. Munro, Editor. 1969, Academic Press. p. 21–132.
26. Vitte C, Chaparro C, Quesneville H, and Panaud O: ***Spip* and *Squiq*, two novel rice non-autonomous LTR retro-element families related to *RIRE3* and *RIRE8*.** *Plant Science*, 2007: 172(1): p. 8-19.
27. Gaut BS, Morton BR, McCaig BC, and Clegg MT: **Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*.** *Proceedings of the National Academy of Sciences of the USA*, 1996: 93: p. 10274-10279.
28. Pelsy F: **Untranslated leader region polymorphism of *Tvv1*, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*.** *Theoretical and Applied Genetics*, 2007: 116(1): p. 15-27.
29. Hudak J and McClure MA: **A comparative analysis of computational motif-detection methods.** *Pac Symp Biocomput*, 1999: p. 138-49.
30. Riccioni C, Rubini A, Belfiori B, Passeri V, Paolocci F, and Arcioni S: ***Tmt1*: the first LTR-retrotransposon from a *Tuber* spp.** *Current Genetics*, 2008: 53(1): p. 23-34.
31. Di Gaspero G, Cipriani G, Adam-Blondon A-F, and Testolin R: **Linkage maps of grapevine displaying the chromosomal locations of 420 microsatellite markers**

and 82 markers for R -gene candidates. *Theoretical and Applied Genetics*, 2007: 114(7): p. 1249-1263.

32. Galet P, **Cépages et vignobles de France. Les vignes américaines**, ed. Edition. 1988. 553.

Eléments	Taille (pb)	LTR	
		Taille (pb)	% Identité
V1	2953	625	99,5
V11	6522	697	100
V12	5433	676	100
V13a	4687	223	100
V13b	4777	268	100
V14	9743	1251	100

Tableau 5 : Caractéristiques structurales des 5 éléments V.

Chapitre II : Les rétrotransposons « V »

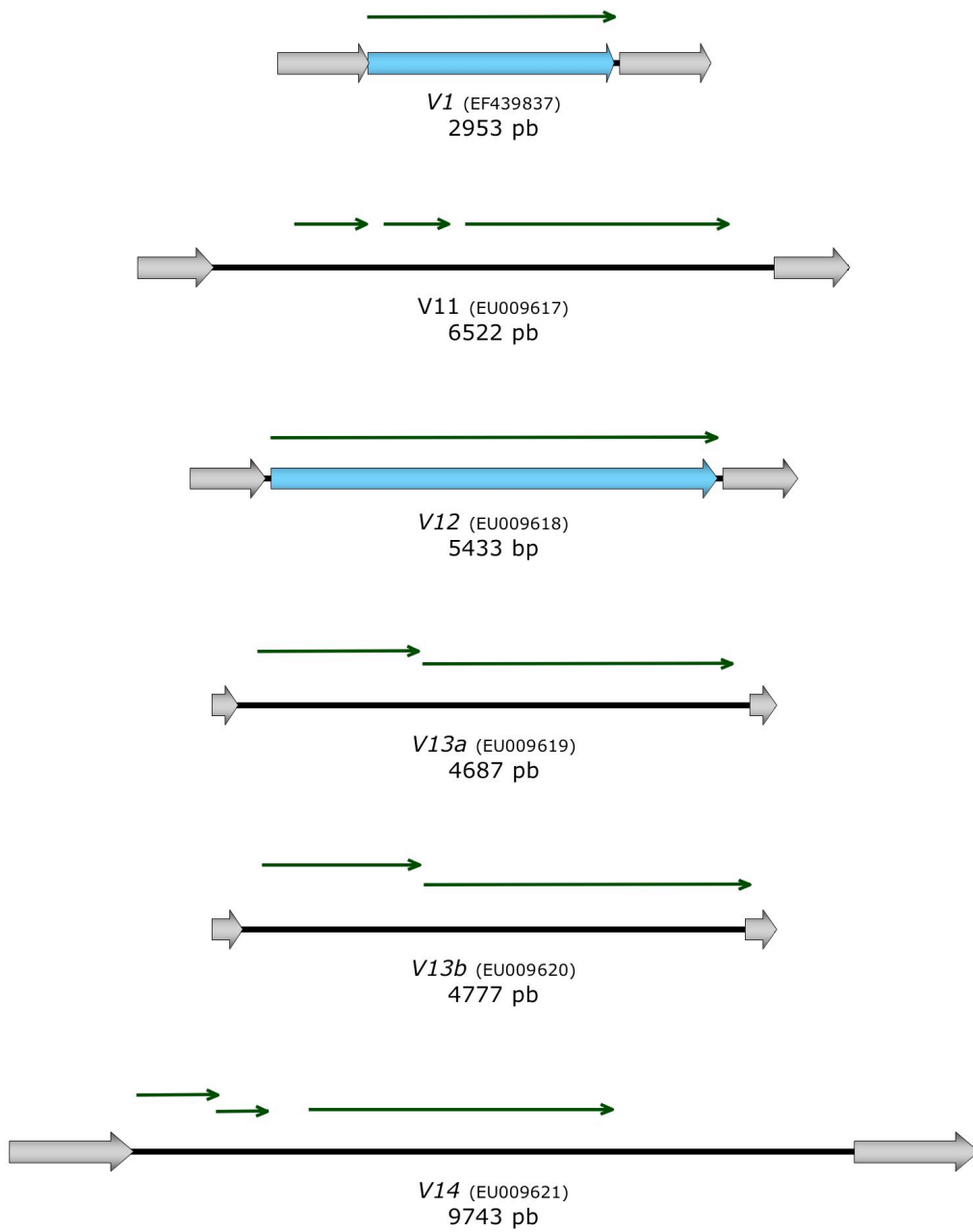
II.1 - Des rétrotransposons de vigne annotés dans les banques de données

En parallèle notre travail sur la caractérisation des 10 nouvelles familles, 8 rétrotransposons à LTR du génome de la vigne ont été annotés par l'équipe du Laboratoire Plant Genomics de l'Université d'Helsinki, en Finlande (données disponibles dans les banques de données NCBI, mais non publiées en 2007). Il s'agit des éléments *VI* (n° accession : EF439837), *V2* (EU009616), *VII* (EU009617), *VI2* (EU009618), *VI3a* (EU009619), *VI3b* (EU009620), *VI4* (EU009621) et *VI6* (EU009622). Il est important de noter qu'aucune indication n'est disponible pour le moment sur la variété de vigne à partir de laquelle ces éléments ont été détectés, ni sur la stratégie utilisée pour les identifier.

En comparant la séquence de ces rétrotransposons à celles des 10 familles que nous avons identifiées, nous avons constaté que les éléments *V2* et *VI6* correspondent respectivement à des éléments des familles *Gans* et *Noble*. De plus, les éléments *VI3a* et *VI3b* appartiennent à la même famille.

II.2 - Prise en compte de 5 familles *V* dans notre étude

A la recherche de données pour étudier l'activité transcriptionnelle des rétrotransposons du génome de la vigne (voir la partie III de ce manuscrit), nous avons décidé d'inclure les éléments *VI*, *VII*, *VI2*, *VI3* et *VI4* à notre étude (Tableau 5). Dans ce but, nous n'avons pas cherché à caractériser toutes les copies de ces familles, ni à délimiter précisément les domaines fonctionnels qui les composent puisque ces recherches ont certainement déjà été entreprises par l'équipe qui les a découverts. Nous avons seulement cherché à déterminer si ces cinq copies sont potentiellement autonomes et combien de copies appartenant à ces familles sont présentes dans le génome de la vigne.



→ = cadre de lecture ouvert, — = ADN non-codant, → = Séquence codant potentiellement pour une polyprotéine *gag-pol* complète, ⇨ = LTR.

Figure 28 : Principaux cadres de lectures ouverts détectés dans la séquence des rétrotransposons V.

a) Recherche de la séquence codante des copies VI, VII, VI2, VI3a, VI3b et VI4

L'identité des deux LTRs de chaque rétrotransposon semble indiquer que ces éléments se sont insérés relativement récemment dans le génome de la vigne (Tableau 5). Nous ne disposons pas de suffisamment de données pour déterminer si ces rétrotransposons *VI*, *VII*, *VI2*, *VI3* et *VI4* appartiennent aux groupes *Ty1-copia* ou *Ty3-gypsy-like*. Il faudrait délimiter précisément les domaines fonctionnels de chaque polyprotéine et se baser sur leur disposition pour conclure à leur appartenance à l'un des deux groupes. Pourtant, en se basant sur le nombre de cadres de lecture ouverts uniques qui codent pour la polyprotéine, il est possible de supposer que l'élément *VI2*, qui contient un cadre de lecture unique et suffisamment grand pour permettre la synthèse d'une polyprotéine complète, pourrait correspondre à une copie autonome du groupe *Ty1-copia-like* (Figure 28).

A l'inverse, les éléments *VI3a*, *VI3b* présentent potentiellement deux cadres de lecture qui pourraient coder pour des protéines *GAG* et *POL* distinctes, une particularité généralement observée chez les rétrotransposons *Ty3-gypsy* (Figure 28). L'hypothèse selon laquelle la présence de deux ORF résulterait de mutations ponctuelles ayant entraîné l'apparition d'un codon stop et d'un second site d'initiation de la transcription, n'est cependant pas exclue, ce qui signifierait que ces éléments sont plus proches du groupe *Ty1-copia*. L'élément *VI* pourrait quant à lui correspondre à une copie tronquée, puisque sa taille (2953 pb) est proche de celle de l'élément *Vine-1* et que ses LTR (625 pb) semblent relativement grandes pour un élément de cette taille. Comme la copie *Vine-1* détectée dans le gène *Adh* de la vigne (Verries *et al.*, 2000), l'élément *VI* contient un cadre de lecture ouvert qui couvre la quasi totalité de sa région interne, qui n'est interrompu par aucun codon stop, mais qui semble trop court pour permettre la synthèse d'une polyprotéine complète (Figure 28). Il est donc possible que cette copie soit capable de synthétiser un ARN couvrant presque toute sa séquence, mais que celui-ci ne contienne pas tous les domaines nécessaires à l'autonomie de la copie. Les éléments *VII* et *VI4* présentent plusieurs cadres de lecture ouverts, mais il est difficile de savoir si cette caractéristique est typique de leur famille ou s'il s'agit de copies dégénérées à partir d'une copie matrice contenant un seul ORF (Figure 28).

b) Evaluation du nombre d'éléments « V » dans le génome de PN40024

En utilisant les mêmes méthodes que précédemment (voir les conditions du Blat dans le chapitre précédent), nous avons recherché les copies potentiellement complètes des cinq rétrotransposons *V* dans le génome de PN40024, en ne retenant que les copies qui présentaient

Elément	n°accession	Nb	% Identité				Taille (pb)				Localisation																				
			min.	med.	moy.	max.	min.	med.	moy.	max.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	U	
V1	EF439837	82	85,5	94,6	93,45	97,8	2374	2619	2639	2946	5	1	2	2	3	4	2	2	7	1	1	4	3	4	2	2	1	5	4	27	
V11	EU009617	21	94,8	96,4	96,17	96,9	5695	6372	6393	7297	1			2	4	1	1					3		1	2			1		5	
V12	EU009618	25	91,9	93,2	93,1	93,8	4431	5322	5404	7642	2		1	3	4		1			2	2			1	1	1		2	1	4	
V13a	EU009619	6	98,2	98,65	98,73	99,3	4165	4695	4613	4779														1	1			1	3		
V14	EU009621	4	94,8	95,7	95,85	97,2	7875	8397	8386	8873												1	1							2	
Total		138									Total	8	1	3	5	9	8	4	3	7	3	3	8	3	8	6	3	1	8	6	41

Taille = taille de la copie de référence, Nb = Nombre de copies présentant une similarité >75% sur plus de 80% de la longueur de la séquence de référence, % Identité = pourcentage d'identité des copies comparées à la copie de référence, min. = minimum, med. = médiane, max. = maximum, Localisation = nombre de copies positionnées sur les chromosomes, U = séquence non-positionnée sur un chromosome

Tableau 6 : Evaluation du nombre de copies complètes des rétrotransposons V dans le génome de PN40024.

au moins 80% de la taille de la copie de référence, et qui montraient au moins 75% d'identité avec celle-ci. Nous avons identifié entre 4 et 82 copies selon la famille, réparties de manière équivalente sur les différents chromosomes du génome de PN40024 (Tableau 6). Selon nos critères de recherche, la famille la plus nombreuse est celle de l'élément *VI*, qui compte 82 copies. Il est intéressant de noter qu'aucune de ces copies ne correspond à un élément de plus grande taille que la copie *VI* annotée, ce qui peut vouloir dire que soit cet élément s'est multiplié malgré le fait qu'il soit probablement non-autonome, soit la ou les copies matrices ont été éliminées lors des cycles d'autofécondation qui ont été nécessaires à l'obtention de l'accession PN40024. La famille la moins représentée est celle de l'élément *VI4* qui ne compte que 4 copies répondant à nos critères de recherche.

En élargissant nos critères aux copies dégénérées de plus petites tailles, nous avons finalement identifié 674 copies correspondant à ces cinq familles. Au total, ces familles représentent approximativement 1,6 Mb, soit environ 0,34% du génome de PN40024.

Les séquences exactes des copies qui ont été annotées ne sont pas présentes dans le génome de PN40024 puisque nous n'avons trouvé aucune copie présentant 100% d'identité et la même taille que les copies *VI*, *VII*, *VI2*, *VI3a* et *VI4*. Ces copies particulières sont peut être spécifiques d'une variété ou ont été éliminées du génome de PN40024 au cours des cycles d'autofécondation.

Type	Famille	Nombre copies	Couverture (Kb)	Taille moy (kb)*	% génome	
Eléments de classe I		33 118	50863,3**	1,536**	10,65	
Protéines détectées par BlastX	Non-LTRs	LINEs	5 504	6 792,9	1,234	1,42
	LTR	<i>Ty1-copia</i>	17 293	24 640,8	1,425	5,16
		<i>Ty3-gypsy</i>	9 632	17 659,6	1,833	3,70
	Autres	Autres	88	103,6	1,177	0,02
		Autres	601	166,6	0,277	0,35
Eléments de classe II		1 797	1 975,9	1,100	0,41	
Hélitrons		109	58,9	0,540	0,01	
Total		35 024	52 898,0	1,510	11,08	
Eléments de classe I		105 532	81 363,7	0,771	17,04	
Annotation manuelle	Non-LTRs	LINEs	15 216	12 131,1	0,797	2,54
	LTR	<i>Ty1-copia</i>	56 890	39 848,3	0,700	8,35
		<i>Ty3-gypsy</i>	14 093	15 339,8	1,088	3,21
	Autres	Autres	18 688	13 191,5	0,706	2,76
		Autres	645	853,0	1,322	0,18
Eléments de classe II		6 344	2 040,9	0,322	0,43	
Hélitrons		0	nc	nc	nc	
Total		111 876	83 404,7	0,746	17,47	

* Estimation de la taille moyenne des copies, d'après les données de Jaillon et al. 2007.

** D'après les données présentées dans Jaillon et al. 2007, la somme des éléments de classe I représenterait en fait 49 364Kb, soit une taille moyenne de 1,491kb et environ 9,87% du génome, .

Tableau 7 : Estimation de la taille moyenne des éléments transposables présents dans le génome de PN40024 et de la proportion des groupes de classe I (d'après Jaillon *et al.* 2007).

Chapitre III : Conclusion

La détection *in silico*, par BlastX et par des annotations manuelles, a permis de détecter toutes les classes et superfamilles d'éléments transposables dans le génome de PN40024 (Jaillon *et al.*, 2007). Au total, 111876 éléments transposables ont été annotés manuellement, représentant plus de 17% du génome. Les résultats présentés dans le Tableau 7 montrent, quelle que soit la méthode de détection utilisée, une très large prédominance des éléments de classe I sur les éléments de classe II et les Hélitrons. Le nombre d'éléments de classe I est supérieur à 105000 alors que moins de 6500 éléments de classe II ont été mis en évidence. En nous basant sur les données obtenues par les deux stratégies d'annotation, nous avons pu évaluer que la taille moyenne des éléments de classe I est de 1,5 et 0,7 kb, contre 1,1 et 0,3 kb pour les éléments de classe II (Tableau 7). Cette estimation peut vouloir dire que les éléments présents dans le génome de la vigne sont relativement petits et/ou dégénérés, même s'il n'est pas impossible que la stratégie d'annotation, par exemple la recherche des régions codant pour les polyprotéines, ait contribué à sous-estimer leur taille ne considérant pas les régions non-codante des éléments, comme les régions LTR ou UTL dans le cas des rétrotransposons.

En raison d'un plus grand nombre de copies, mais également d'une taille moyenne supérieure à celle des transposons, les éléments de classe I représentent également une part nettement plus importante du génome de la vigne (17% contre 0,4%). Les rétrotransposons à LTR représentent 86% de la place occupée dans le génome par les éléments de la classe I (Tableau 7).

Les familles *Gret1*, *Tvv1* et *Vine-1*, les 10 nouvelles familles que nous avons caractérisées, ainsi que les 5 éléments « V », représentent donc les 18 familles de rétrotransposons du génome de la vigne connues à ce jour. D'après nos résultats, ces 18 familles représentent plus de 2380 copies et couvrent environ 6,4 Mb du génome, c'est-à-dire qu'elles représentent, selon la méthode d'annotation considérée, 8,8% ou 4,2% des copies de rétrotransposons à LTR du génome, et 15% ou 11,6% de la part du génome occupée par les rétrotransposons à LTR (d'après les données de Jaillon *et al.*, 2007). Certaines familles semblent s'être multipliées récemment, et quelques copies potentiellement actives et autonomes ont été détectées dans le génome de PN40024.

Bien que non-exhaustives, ces nouvelles données nous ont permis d'envisager l'étude de l'activité des rétrotransposons à une échelle plus importante au fur et à mesure de l'acquisition des données.

**Partie III : Étude de l'activité
transcriptionnelle des
rétrotransposons**

Chapitre I : Recherche de rétrotransposons dans les banques EST de vigne

I.1 - Blast sur les banques de séquences EST de vigne

La première stratégie utilisée pour évaluer le niveau de transcription des rétrotransposons a consisté à rechercher des séquences spécifiques de chaque famille de rétrotransposons dans les banques de séquences exprimées disponibles pour la vigne. Cette recherche a été réalisée par Blast (option MEGABLAST) sur le site NCBI (www.ncbi.nlm.nih.gov/BLAST/) en utilisant l'option « *Vitis* : taxid n°3603 ». Le nombre exact de séquences EST correspondant au genre *Vitis* n'était pas disponible. Cependant, lorsque j'ai réalisé cette recherche, 336532 séquences EST de *Vitis vinifera* étaient réparties dans des banques générées à partir plusieurs variétés de vignes (Cabernet sauvignon, Muscat d'Hambourg, Pinot noir...), sur différents organes (baies, feuilles, bourgeons, racines...) et dans des conditions physiologiques différentes (croissance normale, stress hydrique, stress abiotique, stress chimique...). Il est important de noter qu'aucune indication n'est disponible sur le nombre de séquences EST contenu dans chaque banque, et qu'il nous est donc impossible d'évaluer leur contribution au nombre total de séquences EST disponibles dans la base de données. De plus, les banques représentées dans cette base de données ont été construites par des équipes indépendantes, qui ont utilisées des techniques de clonage et de séquençage différentes. Les séquences EST disponibles dans cette base de donnée présentent donc des tailles très hétérogènes.

Pour ne privilégier aucune de ces conditions, nous n'avons pas fixé de limite minimale de taille et toutes les séquences EST qui présentaient au moins 80% d'identité avec la copie requête et une e-value supérieure à e^{-150} , ont été conservées pour l'analyse. Lorsqu'ils ont recherché les EST de rétrotransposons chez les Graminées, Vicient et collaborateurs ont pu utiliser la recherche par mots clefs « *copia* » ou « *gypsy* » (Vicient *et al.*, 2001), ce qui n'a pas été possible dans notre cas puisque la majorité des séquences contenues dans les banques EST de vigne n'est pas annotée, nous aurions donc éliminé trop de données de notre étude.

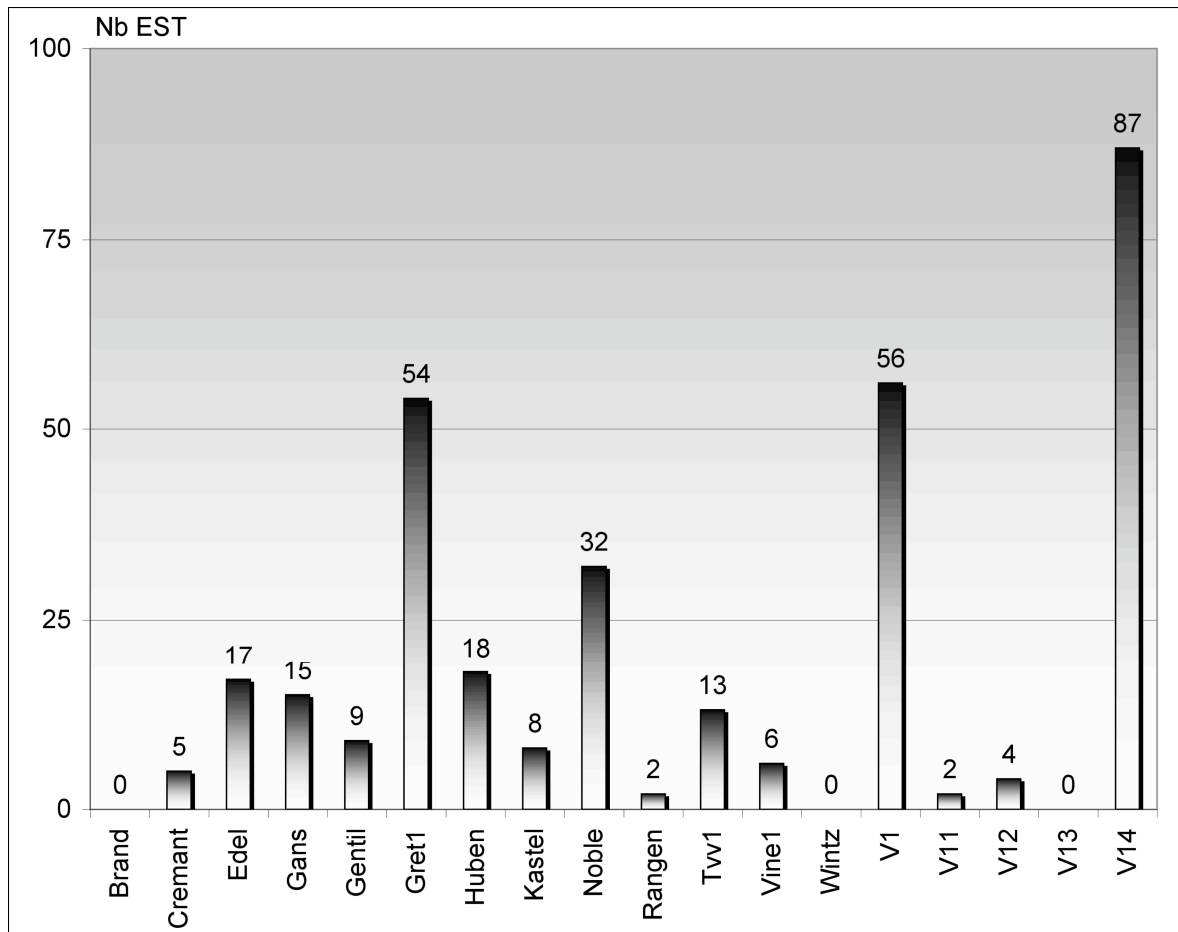


Figure 29 : Séquences EST trouvées dans les bases de données *Vitis* pour les 18 familles de rétrotransposons.

I.2 - Choix des séquences requêtes

La séquence nucléique complète d'une copie de chacune des 18 familles de rétrotransposons a été utilisée comme requête pour la recherche par Blast . Lorsque cette étude a été réalisée, toutes les copies de ces familles n'avaient pas été encore identifiées. Nous avons donc choisi, parmi celles dont nous disposions, la séquence de l'élément qui présentait les caractéristiques les plus proches de celles attendues chez une copie active, c'est-à-dire : des LTRs 5' et 3' identiques à 100% ou presque (preuve que la copie s'est insérée récemment), et un cadre de lecture ouvert supposé complet et non-interrompu par un codon stop (Tableau 9).

I.3 - Résultats

a) Représentativité des 18 familles

La recherche par Blast dans les banques de données de *Vitis* nous a permis d'identifier 328 séquences EST présentant des similarités avec 15 copies requêtes parmi les 18 utilisées (Figure 29). Les familles de rétrotransposons peuvent classées en quatre catégories, selon le nombre d'EST leur correspondant dans les banques : celles pour lesquelles aucune séquence EST n'a été détectée : les familles *Brand*, *Wintz* et *VI3* ; celles pour lesquelles entre 2 et 10 séquences ont été trouvées : les familles *Crémant*, *Gentil*, *Kastel*, *Rangen*, *Vine-1*, *VII* et *VI2* ; celles pour lesquelles entre 10 et 20 séquences EST ont été détectées : les familles *Edel*, *Gans*, *Huben* et *Tvv1* ; et enfin, les familles les plus représentées : *Gret1*, *Noble*, *VI* et *VI4* avec respectivement 54, 32, 56 et 87 séquences EST.

b) EST détectées dans plusieurs variétés

Les EST de rétrotransposons appartiennent à 31 banques qui ont été obtenues dans des conditions physiologiques différentes, sur des vignes cultivées (16 banques ont été réalisées à partir de la variété Cabernet sauvignon). Parmi les 328 séquences EST de rétrotransposon détectées dans cette étude, 217 (66,2%) proviennent de banques construites à partir de la variété Chardonnay et 75 (22,9%) à partir du Cabernet sauvignon (Figure 30). Les 36 séquences EST restantes (10,9%) ont été détectées dans des banques obtenues à partir de 9 autres variétés de vigne (Moser *et al.*, 2005; Peng *et al.*, 2007) (ou données non-publiées).

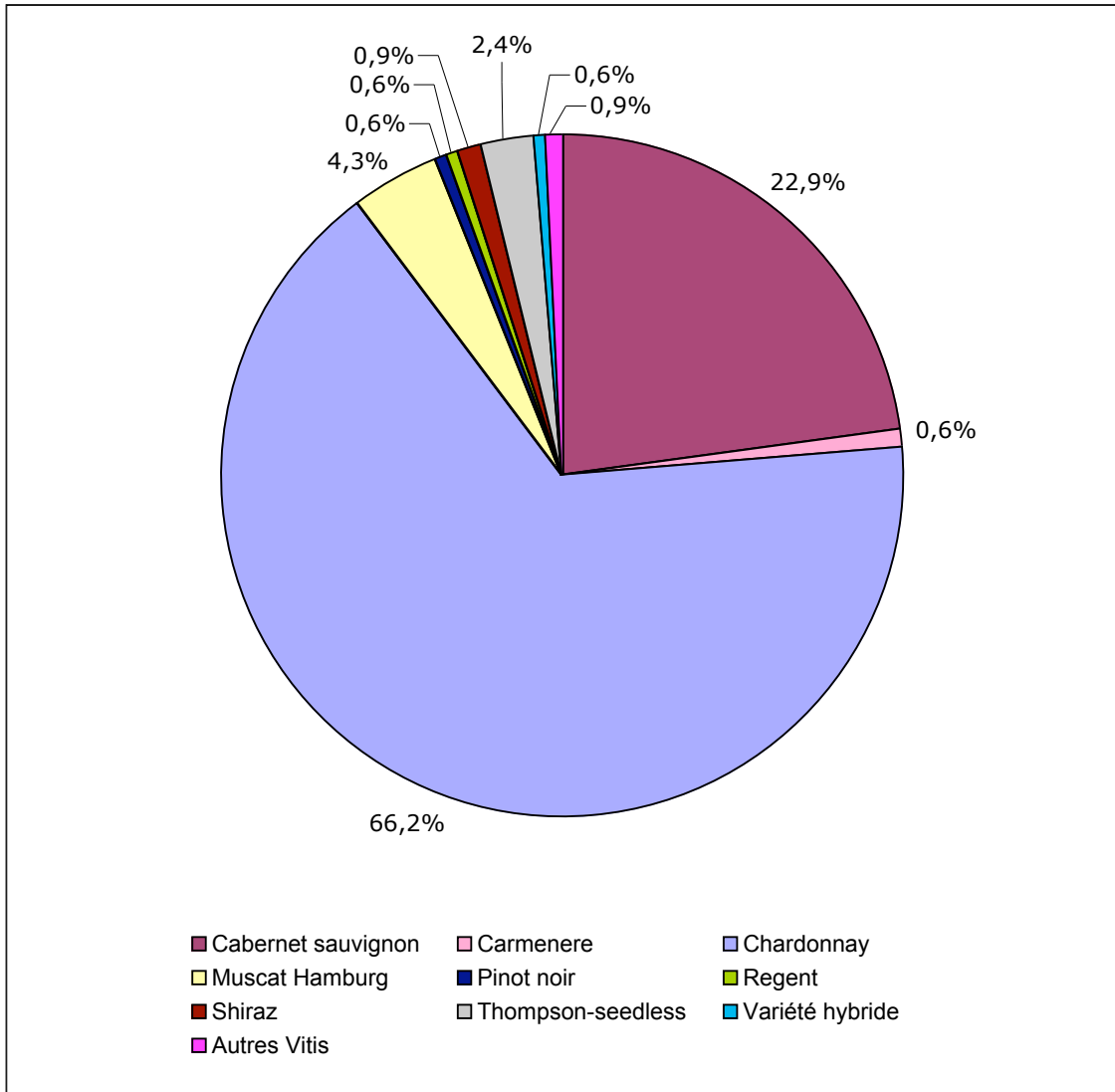


Figure 30 : Répartition des EST de rétrotransposons parmi les genres et variétés de vigne utilisés pour la construction des banques.

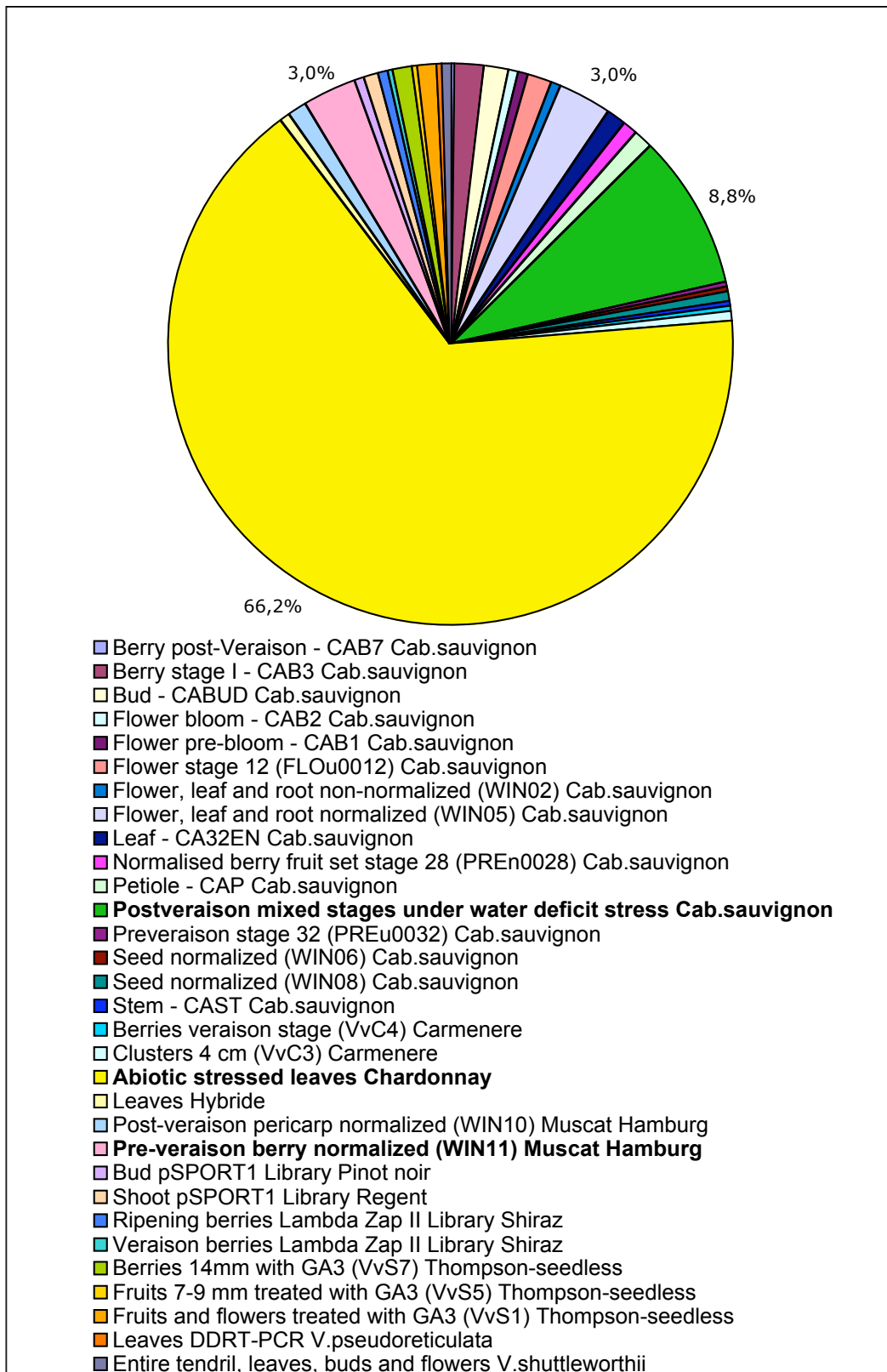


Figure 31 : Répartition des séquences EST de rétrotransposons dans les 31 conditions.

Description de la banque EST	Variétés	Références	Rétrotransposons																					
			Brand	Cremant	Edel	Gans	Gentil	Gret1	Hubben	Kastel	Noble	Rangen	Tw1	Vine1	Wintz	V1	V11	V12	V13	V14	Total			
Berry post-veraison - CAB7	Cab.sauvignon	unpublished													1								1	
Berry stage I - CAB3	Cab.sauvignon	unpublished																						5
Bud - CABUD	Cab.sauvignon	unpublished								2						2								5
Flower bloom - CAB2	Cab.sauvignon	unpublished								1											1			2
Flower pre-bloom - CAB1	Cab.sauvignon	unpublished																			1			2
Flower stage 12 (FLOu0012)	Cab.sauvignon	unpublished								1						1								4
Flower, leaf and root non-normalized (WIN02)	Cab.sauvignon	Peng et al. (2007)														1								2
Flower, leaf and root normalized (WIN05)	Cab.sauvignon	Peng et al. (2007)								1						1								4
Leaf - CA32EN	Cab.sauvignon	unpublished							4															4
Normalised berry fruit set stage 28 (PREn0028)	Cab.sauvignon	unpublished								2						1								2
Petiole - CAP	Cab.sauvignon	unpublished								2						2								4
Postveraison mixed stages under water deficit stress	Cab.sauvignon	unpublished								8						1								29
Preveraison stage 32 (PREu0032)	Cab.sauvignon	unpublished																						1
Seed normalized (WIN06)	Cab.sauvignon	Peng et al. (2007)								1														1
Seed normalized (WIN08)	Cab.sauvignon	Peng et al. (2007)														1								2
Stem - CAST	Cab.sauvignon	unpublished														1								1
Berries veraison stage (VvC4)	Camenera	unpublished																						1
Clusters 4 cm (VvC3)	Camenera	unpublished														1								1
Abiotic stressed leaves	Chardonay	unpublished	5	7	7	3	44	12	2	14	2	4	46	1	70	217								
Leaves	Hybride	unpublished					1																	2
Post-veraison pericarp normalized (WIN10)	Muscat Hamburg	Peng et al. (2007)									2					1								4
Pre-veraison berry normalized (WIN11)	Muscat Hamburg	Peng et al. (2007)								1						1								3
Bud pSPORT1 Library	Pinot noir	Moser et al.(2005)								1														2
Shoot pSPORT1 Library	Regent	Moser et al.(2005)								1														2
Ripening berries Lambda Zap II Library	Shiraz	unpublished								2														2
Veraison berries Lambda Zap II Library	Shiraz	unpublished								1														1
Berries 14mm with GA3 (VvS7)	Thompson-seedless	unpublished								4														4
Fruits 7-9 mm treated with GA3 (VvS5)	Thompson-seedless	unpublished																			1			1
Fruits and flowers treated with GA3 (VvS1)	Thompson-seedless	unpublished								1														3
Leaves DDRT-PCR	<i>V.pseudoreticulata</i>	unpublished																						1
Entire tendril, leaves, buds and flowers	<i>V.shuffleworthii</i>	unpublished								1														2
			Total	0	5	17	15	9	54	18	8	32	2	13	6	0	56	2	4	0	87	328		

Tableau 8 : Nombre de séquences EST obtenues par condition pour chaque rétrotransposon.

En choisissant de comparer les séquences requêtes à toutes les séquences EST disponibles pour le genre *Vitis* (option du Blast), nous avons pu identifier une séquence EST de l'élément *VI2* dans une banque construite à partir de *Vitis pseudoreticulata* (vigne asiatique) et deux séquences EST de *Gentil* et *Kastel* dans une banque obtenue à partir de *Vitis shuttleworthii* (vigne américaine) (Tableau 8).

c) EST de rétrotransposons plus nombreuses en conditions de stress

Alors que des séquences EST de rétrotransposons ont été détectées dans 31 conditions différentes, les EST correspondant aux familles *Crémant*, *Rangen*, *Vine-1* et *VII* n'ont été détectées que dans 1, 2, 3 et 2 conditions respectivement (Tableau 8). Au contraire, les séquences EST des 14 autres éléments ont été trouvées dans 4 à 8 conditions, et la famille *Noble* a été détectée dans 10 conditions différentes.

Au total, 238 (72,6%) séquences EST de rétrotransposons ont été détectées dans des banques construites à partir de feuilles, 30 (9,2%) à partir de baies et 60 (18,2%) à partir d'autres organes (Tableau 8).

Plus particulièrement, 217 (66,2%) séquences EST correspondant à 13 familles ont été obtenues à partir de feuilles de Chardonnay soumises à une condition de stress abiotique, dont la nature n'est malheureusement pas précisée, ce qui représente 91,2% des séquences EST détectées dans les banques construites à partir de feuilles (Figure 31). Toutes les séquences EST correspondant à l'élément *Crémant* ont été détectées dans cette unique condition, ainsi que plus de 80% des séquences EST des éléments *Gret1*, *VI* et *VI4* (Tableau 8). Par exemple, 44, 12, 14 et 70 EST correspondent respectivement aux familles *Gret1*, *Huben*, *Noble* et *VI4* dans cette condition, alors que ces familles ne comptent que 4, 1, 0 et 2 EST dans les autres banques construites à partir de feuilles. De la même manière, l'élément *VI* correspond à 46 EST de cette banque mais à aucune dans les autres banques construites à partir de feuilles.

Par ailleurs, 29 (8,8%) séquences EST de rétrotransposons ont été détectées dans une banque construite à partir de baies de Cabernet sauvignon soumises à un déficit hydrique, dont la nature n'est malheureusement pas précisée (Figure 31).

Finalement, 75% des séquences EST de rétrotransposons ont été identifiées dans ces deux banques, qui correspondent à des variétés de vigne soumises à des conditions de stress, et plus particulièrement 88% des séquences EST de *Edel*, 85% de *Gret1*, 86% de *VI* et 90% de *VI4*, qui sont parmi les éléments les plus représentés dans les banques EST.

Requêtes						
Famille	Copie ou n° accession	Taille (pb)		nb EST		
		Séquence complète	LTR	LTR (chimères)	Interne	Total
<i>Brand</i>	Brand-AF01	5519	501-502	0	0	0
<i>Crémant</i>	Crement-AF01	4807	244	0	5	5
<i>Edel</i>	Edel-BO4	4783	376	2 (1)	15	17
<i>Gans</i>	Gans-B01	4217	206	13 (4)	2	15
<i>Gentil</i>	Gentil-BO1	4265	320-295	8	1	9
<i>Gret1</i>	AB111100	10422	824	8	46	54
<i>Huben</i>	KG118-7	4941	206	4	14	18
<i>Kastel</i>	KGii-6-3-2	4072	298	7	1	8
<i>Noble</i>	Noble-AF16	5502	201	11 (7)	21	32
<i>Rangen</i>	KGii-1-3	5101	149	0	2	2
<i>Tvv1</i>	AF478367	5222	149	0	13	13
<i>Vine1</i>	AF116598	2392	287	4	2	6
<i>Wintz</i>	KG116-7	5278	291	0	0	0
<i>V1</i>	EF439837	2953	625	12	44	56
<i>V11</i>	EU009617	6522	697	2	0	2
<i>V12</i>	EU009618	5433	676	2	2	4
<i>V13</i>	EU009619	4687	223	0	0	0
<i>V14</i>	EU009621	9743	1251	59	28	87
Total				134	194	328

Requêtes : rétrotransposon dont la séquence nucléique complète a été utilisée comme modèle pour la recherche par BLASTN. Taille : taille de la séquence requête. Nb EST : nombre de séquences EST correspondant aux différents domaines de la requête. LTR : no

Tableau 9 : Position des séquences EST sur les séquences requêtes.

d) Position des EST sur les séquences requêtes

L'alignement par ClustalW des séquences EST avec la séquence requête a révélé que les EST identifiées ne correspondent pas toutes à la même région du rétrotransposon (Tableau 9). Sur les 328 séquences EST détectées, 194 (59,1%) correspondent à la région interne des éléments requêtes, 122 (37,2%) correspondent soit entièrement aux LTRs, soit sont à cheval entre un domaine LTR et la région interne du rétrotransposon, et enfin 12 (3,7%) sont des séquences chimériques, comprenant une partie du rétrotransposon et une partie indéterminée.

Trois types de répartitions ont été observés, illustrés par les trois exemples présentés dans la Figure 32 : dans le cas de l'élément *Crémant*, les séquences EST appartiennent toutes à la région interne ; dans le cas de l'élément *Gretl*, les EST ne sont pas réparties de façon uniforme sur la séquence, la majorité d'entre elles correspondant au domaine ORF 2 ; et dans le cas de l'élément *Noble*, les séquences EST sont réparties sur toute la longueur de la séquence requête. Sur 11 séquences EST positionnées dans les LTRs de *Noble*, 7 séquences présentaient des extrémités 5' ou 3' n'ayant aucune similarité avec la séquence requête et ont par conséquent été considérées comme chimériques. Les 12 séquences EST chimériques détectées dans cette étude correspondent aux trois familles *Edel*, *Gans* et *Noble* (Tableau 9). Les extrémités qui ne correspondent pas aux séquences des rétrotransposons requêtes sont trop courtes pour déterminer si elles appartiennent effectivement à la séquence de l'hôte.

Dans le cas des éléments *Crémant*, *Rangen* et *Tvv1*, aucune séquence EST ne correspond aux régions LTR alors que 5, 2 et 13 EST correspondent à leurs régions internes respectives (Tableau 9). À l'inverse, aucune séquence EST située dans la région interne de l'élément *VII* n'a été détectée, alors que 2 EST ont été trouvées dans ses LTRs. Dans le cas de l'élément *VI4*, 59 (67,8%) séquences EST correspondent à la région LTR et 28 (32,2%) à la région interne.

Pour les deux rétrotransposons Ty3-gypsy-like, qui sont deux éléments de grande taille, 141 (43%) séquences EST ont été détectées : 54 EST pour *Gretl* (10422 pb) et 87 EST pour *VI4* (9743 pb) (Tableau 9). L'ensemble des rétrotransposons Ty1-copia-like représentent 187 (57%) séquences EST. Parmi eux, l'élément *VII* (6522 pb), qui est l'élément le plus grand, ne correspond qu'à 2 séquences EST, alors que *VI* (2953 pb), l'élément le plus petit, est représenté par 56 séquences EST. Sur les 18 séquences requêtes, 13 ont des tailles comprises entre 4072 et 5520 pb. Parmi ces 13 rétrotransposons, 8 séquences EST ont été trouvées pour

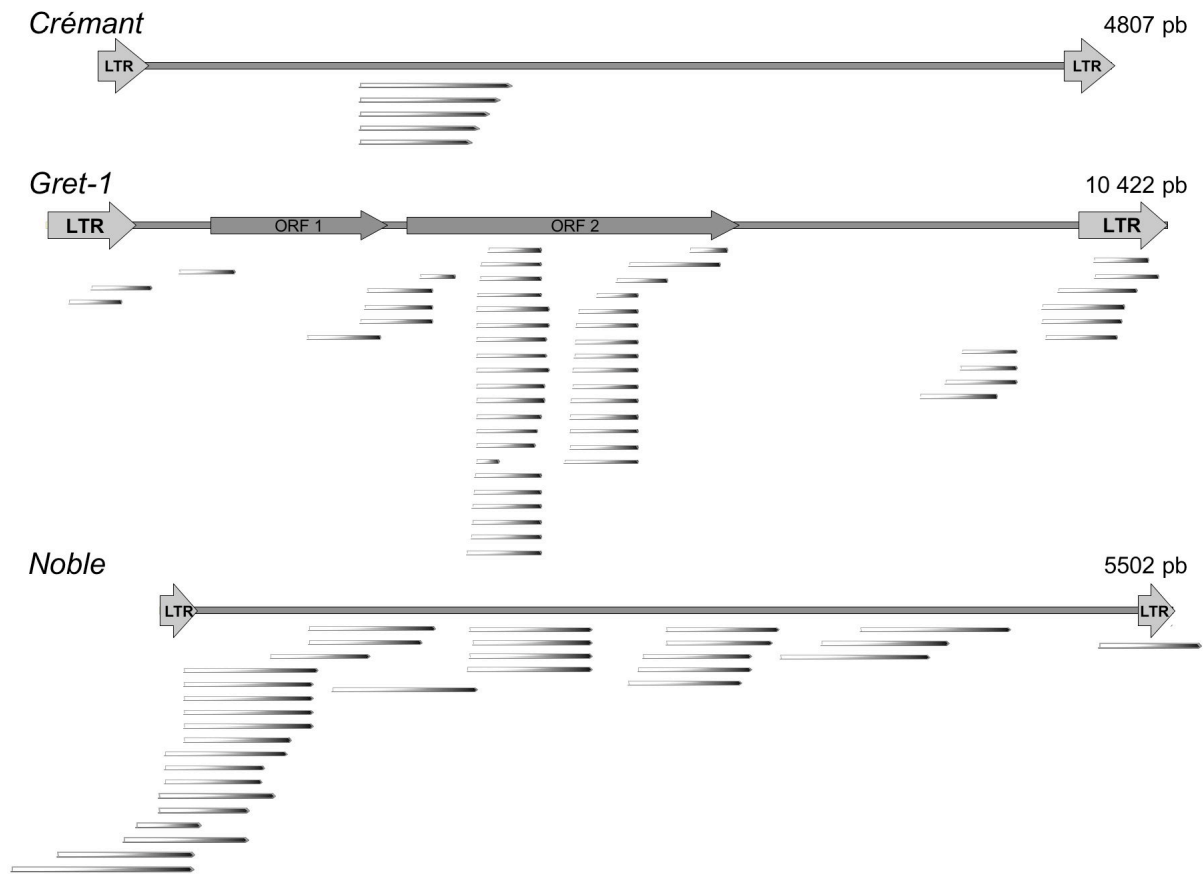


Figure 32 : Localisation des séquences EST sur la séquence requête : Exemples des rétrotransposons *Crémant*, *Gret1* et *Noble*.

l'élément le plus petit, *Kastel*, alors qu'aucune séquence EST n'a été détectée pour l'élément le plus grand, *Brand*.

I.4 - Discussion

a) Transcription des rétrotransposons chez la vigne

Des séquences de rétrotransposons sont présentes dans les banques de données EST disponibles pour la vigne. Au total, 328 séquences EST présentant des similarités significatives avec 15 familles de rétrotransposons ont été trouvées par Blast, en utilisant comme requêtes les séquences canoniques des 18 familles connues du génome de la vigne. Parmi ces séquences, 325 EST ont été obtenues sur des variétés de l'espèce *V. vinifera*, pour laquelle 336532 séquences EST étaient disponibles dans les banques. Au total, les séquences EST des 18 familles de rétrotransposons représentent environ 0,97‰ des EST disponibles pour *V. vinifera*. D'après Boroluzzi et collaborateurs, les transcrits peuvent être considérés comme abondants lorsqu'ils représentent plus de 3,6‰ des EST disponibles, modérés entre 3,6‰ et 1,25‰, et rares en dessous de 1,2‰ (Bortoluzzi et al. 2000). Selon ces critères, les séquences EST des familles de rétrotransposons étudiées sont rares chez la vigne.

Cependant, ce résultat doit être pondéré pour plusieurs raisons.

Tout d'abord, les 18 familles que nous avons étudiées ne représentent pas la totalité des rétrotransposons du génome de la vigne. D'après les données présentées dans la Partie II de ce manuscrit, ces 18 familles représentent plus de 2380 copies et couvrent environ 6,4 Mb du génome, ce qui représente, 8,8% ou 4,2% du nombre copies et 15% ou 11,6% de la part du génome occupée par les rétrotransposons à LTR, selon la méthode d'annotation considérée (d'après les données de Jaillon *et al.*, 2007). Ainsi, la totalité des rétrotransposons à LTR du génome de la vigne représente probablement une part plus importante des séquences EST de vigne. Il est possible que des familles encore inconnues soient beaucoup plus transcrites que celles que nous avons étudiées. Et si le nombre de séquences EST de rétrotransposon est proportionnel au nombre de copies dans le génome, il est possible que les séquences EST de rétrotransposons représentent 6 à 23 fois plus d'EST que celles que nous avons détectées, c'est à dire entre 5,8‰ et 22,3‰ des EST de vigne. Si cette estimation est exacte, cela signifierait, selon les critères proposés par Boroluzzi, que les rétrotransposons sont des transcrits abondants dans les cellules de vigne.

Ensuite, chaque famille de rétrotransposons compte peut être des copies transcrites dont les séquences sont moins similaires à la séquence requête, et pour lesquelles les séquences EST correspondantes n'ont pas été détectées dans cette étude. Ce biais peut nous avoir conduit à sous-évaluer le nombre d'EST de rétrotransposons présent dans les bases de données. Cependant, il est probable que les transcrits qui participent réellement au cycle de rétrotransposition aient des séquences proches de celles des copies actives. Les copies non-autonomes doivent, par exemple, avoir conservé un minimum de domaines spécifiques de la famille, comme les LTRs, pour pouvoir réaliser leur cycle en utilisant la machinerie enzymatique des copies autonomes. Ainsi, les séquences requêtes ayant été choisies pour cibler les copies potentiellement autonomes des familles et les paramètres du Blast étant peu sélectifs, les EST correspondant aux copies actives de chacune des 18 familles recherchées ont probablement été détectées dans cette étude.

A l'inverse, le nombre de séquences EST de rétrotransposons a pu être sur-évalué car la plupart des conditions à partir desquelles les banques ont été réalisées sont des conditions de stress, susceptibles d'induire la surexpression des rétrotransposons. De plus, il est peu probable que chaque banque contribue de façon équivalente à la quantité totale de séquences EST disponibles pour la vigne. En effet, le nombre de clones séquencés pour la construction d'une banque diffère souvent d'une expérience à l'autre. Ainsi, la proportion de séquences EST d'un rétrotransposon trouvée dans une banque dépend de son niveau de transcription dans cette condition, mais également du nombre de séquences EST contenues dans cette banque. A un niveau supérieur, la représentativité d'un rétrotransposon dans la totalité des banques disponibles est également influencée par la contribution de chaque banque à la quantité totale de séquences disponibles dans les bases de données. Par exemple, si une première banque construite dans la condition X contient trois fois plus de séquences EST qu'une seconde banque obtenue dans la condition Y, le nombre d'EST détectées pour un rétrotransposon sera trois fois plus important dans la conditions X que dans la condition Y. De la même façon, deux rétrotransposons dont les niveaux de transcription sont équivalents, paraîtront différemment exprimés s'ils sont détectés dans la condition X ou Y. La représentativité des rétrotransposons dans les bases de données a ainsi pu être sur-estimée du fait de la contribution plus importante d'une banque à la base de données.

b) Un niveau d'expression plus fort que chez d'autres plantes ?

Vicient et collaborateurs ont montré que les éléments transposables représentent en moyenne 1,2 séquences sur 1000 EST chez les plantes qu'ils ont testées : les EST de rétrotransposons représentent 0,85‰ des séquences EST disponibles pour *Solanum*, 0,98‰ pour le riz, 1,25‰ pour le soja, 2,49‰ pour le maïs et 23,8‰ pour le tabac (Vicient *et al.*, 2001). Chez la canne à sucre, 0,31‰ (82) EST correspondant à des rétrotransposons de type Ty1-*copia* et 0,07‰ (19) EST correspondant à des rétrotransposons Ty3-*gypsy* ont été trouvées dans 37 banques, qui représentent au total 260781 séquences (Rossi *et al.*, 2001). Ainsi, le nombre d'EST disponibles et la proportion d'EST d'éléments transposables et de rétrotransposons varient énormément d'une espèce à l'autre.

La proportion de séquences EST correspondant aux 18 familles chez la vigne (0,97‰) est donc proche des résultats obtenus chez d'autres plantes, notamment chez le riz. Compte tenu du fait qu'elle ne représente qu'une partie des rétrotransposons du génome et que des familles actives sont probablement encore inconnues, la vigne pourrait faire partie des espèces végétales dont les rétrotransposons s'expriment le plus fortement.

c) Conditions d'expression

Les 328 séquences EST de rétrotransposons ont été obtenues à partir de 31 banques réalisées dans des conditions différentes et sur 11 variétés de vigne. Bien que plus de la moitié des banques d'EST aient été obtenues à partir de la variété Cabernet sauvignon, celle-ci ne représente que 23% des séquences EST de rétrotransposons. A l'inverse, 66,2% des EST ont été caractérisées dans une banque obtenue à partir de feuilles de la variété Chardonnay soumises à un stress abiotique. Ainsi, toutes les EST des familles *Crémant* et *VI*, et plus de 80% des séquences EST des familles *Gret1* et *VI4* ont été détectées dans cette banque. De plus, les séquences EST correspondant à plusieurs familles sont sensiblement plus nombreuses dans cette condition que dans toutes les autres banques obtenues sur feuilles. Le stress abiotique est probablement à l'origine de la plus forte représentation de ces éléments dans cette banque, mais il est également possible que le génome de la variété Chardonnay contienne plus de copies actives ou que celles-ci s'expriment plus fortement que dans les autres variétés. En conséquence, les séquences EST de rétrotransposons pourraient être sur-représentées dans cette banque sans que le stress appliqué n'ait pour autant d'influence directe sur leur niveau de transcription. Par ailleurs, la banque obtenue à partir de baies de Cabernet sauvignon soumises à un déficit hydrique représente 8% des séquences EST de

rétrotransposons, et 5 familles y sont plus représentées que dans toutes les autres banques obtenues sur des baies. Le nombre important de séquences EST de rétrotransposons détectées dans ces deux banques suggère que ces deux conditions de stress participent à l'induction de la transcription de plusieurs rétrotransposons du génome de la vigne, notamment pour les familles *Crémant*, *Gret1*, *VI* et *VI4*.

Au total, 72,6% des séquences EST de rétrotransposons ont été détectées dans banques EST construites à partir de feuilles, 9,2% à partir de baies et 18,2% à partir d'autres organes de la vigne. Chez les graminées, les EST de rétrotransposons représentent 26,5% des EST détectées dans banques construites à partir de feuilles, 2% dans les racines et les jeunes pousses, 30% dans les fleurs et les autres appareils reproducteurs, 11,6% dans les embryons et les graines, 20% dans les graines (Vicient *et al.*, 2001). S'il est possible que les éléments transposables ne présentent pas les mêmes niveaux d'expression selon la variété de vigne, les taux de transcription des rétrotransposons détectés dans cette étude semblent essentiellement influencés par les conditions dans lesquelles la banque EST a été construite, ce qui pourrait vouloir dire que leur expression est soumise à une régulation spécifique selon le tissu.

Finalement, la détection d'EST de rétrotransposons dans deux espèces américaines et asiatiques du genre *Vitis*, dans des feuilles de *V. pseudoreticulata* infectées par le champignon *Uncinula necator* (Wang et Zhang, 2003, non publié) et dans *V. shuttleworthii* (Leong, 2004, non publié), indique que des rétrotransposons sont probablement transcrits dans d'autres espèces que la vigne cultivée traditionnellement, notamment en réponse à l'attaque d'un pathogène.

d) Représentativité des 18 familles

Les résultats montrent qu'aucune séquence EST ne correspondant aux éléments *Brand*, *Wintz* et *VI3* n'a été identifiée, ce qui pourrait résulter soit du fait que les conditions dans lesquelles ces banques ont été construites n'étaient pas favorables à leur expression, soit que ces trois familles ne présentent aucune copie active dans le génome des variétés à partir desquelles les banques ont été réalisées. Ces résultats sont en accord avec le fait que, dans le génome de PN40024, aucune copie des familles *Brand* et *Wintz* ne présente les caractéristiques d'un élément actif et autonome. Dans le cas de *VI3*, nous avons vu précédemment que les cadres de lecture ouverts des copies *VI3a* et *VI3b* sont interrompus par un codon stop.

Les éléments *Crémant*, *Gentil*, *Kastel*, *Rangen*, *Vine-1*, *V11* et *V12* sont relativement peu représentés dans les banques de séquences EST. Par exemple, des séquences EST correspondant aux familles *Rangen*, *Vine-1* et *V11* sont détectées, en petites quantités, dans moins de quatre conditions. Le niveau de transcription de ces sept familles est probablement faible, ou induit uniquement dans certaines conditions. Dans le cas de la famille *Crémant*, une copie potentiellement autonome est présente dans le génome de PN40024, ce qui peut vouloir dire que cette famille est probablement capable de réaliser son cycle de transposition. A l'inverse, aucune copie potentiellement autonome n'a été détectée pour les 6 autres familles dans le génome de PN40024, indiquant que les copies transcrites ne sont pas autonomes ou que les copies autonomes ont été éliminées lors des cycles d'autofécondation.

Il est intéressant de noter la présence de l'élément *Vine-1* dans les banques d'EST, alors que le premier élément décrit est tronqué, dépourvu de son domaine RT-RNaseH (Verries *et al.*, 2000) et qu'aucune copie complète de la famille *Vine-1* n'a été trouvée jusqu'à présent dans le génome de la vigne, que ce soit par PCR (données personnelles) ou dans la séquence du génome de PN40024 (voir Partie II). Ainsi, la détection de séquences EST tend à prouver que cette famille est effectivement transcrite, mais elle ne signifie pas pour autant que *Vine-1* soit capable de réaliser son cycle de rétrotransposition de manière autonome.

Les rétrotransposons *Edel*, *Gans*, *Gret1*, *Huben*, *Noble*, *Tvv1*, *V1* et *V14* sont les plus représentés dans les banques de séquences EST. Plusieurs EST de ces éléments ont été détectées dans au moins quatre conditions différentes. Ces huit familles représentent presque 90% des séquences EST de rétrotransposons et semblent correspondre aux familles les plus actives au niveau transcriptionnel dans ces conditions. Par exemple, des séquences EST correspondant à la famille *Noble* ont été détectées dans 10 conditions différentes, ce qui peut vouloir dire qu'une (ou plusieurs copies) de cette famille est(sont) transcrite(s) de manière relativement constitutive. Il est intéressant de noter que des copies potentiellement autonomes appartenant aux familles *Edel* et *Noble* ont été détectées dans le génome de PN40024, laissant supposer que ces deux familles sont actives et mobiles chez la vigne.

e) Lien entre la taille des séquences requêtes et le nombre d'EST détectées ?

Théoriquement, la taille et la séquence de l'élément utilisé comme requête peuvent avoir influencé le nombre de séquences obtenues dans cette étude. En effet, plus la séquence requête est grande, et plus il y a statistiquement de chances de trouver des séquences correspondantes par Blast, comme c'est le cas pour les éléments Ty3-gypsy de grandes tailles, *Gret1* (10422 pb) et *VI4* (9743 pb), pour lesquels ont effectivement été trouvées plus de séquences EST. A l'inverse, peu de séquences EST ont été détectées pour l'élément *VII* malgré sa taille importante (6522 pb) alors que l'élément *VI*, dont la taille est parmi les plus petites (2953 pb), est l'un des plus représentés dans les banques. Sur 18 séquences requêtes, 13 avaient des tailles relativement proches mais le nombre de séquences EST trouvées pour ces familles est très variable et sans relation directe avec leur taille respective. Ainsi, le biais occasionné par la différence de taille des séquences requêtes ne semble pas influencer significativement le nombre de séquences EST détectées, qui semble plutôt refléter le niveau d'expression de ces familles dans les conditions particulières de construction de chaque banque.

f) Position des EST et séquences chimériques

L'alignement par ClustalW de la séquence requête et des séquences EST a permis de positionner celles-ci sur les régions du rétrotransposon. Ces alignements ont montré que les EST n'étaient pas réparties de manière uniforme sur tous les domaines, puisque 59% d'entre elles correspondent à la région interne des éléments requêtes, contre 41% aux LTRs. Pourtant, les deux domaines LTRs réunis ne représentent en moyenne que 15% de la taille totale d'un rétrotransposon (moyenne calculée pour les 18 rétrotransposons). Les régions internes étant donc plus grandes et leurs séquences moins variables au sein d'une famille que les domaines LTRs, un écart plus important dans la répartition des EST pouvait être attendu entre ces deux régions. Plusieurs explications sont envisageables pour expliquer la répartition hétérogène des EST sur toute la longueur de la séquence requête.

Origine de la séquence et taille de l'élément

Tout d'abord, les séquences EST appartiennent à des banques distinctes et de ce fait leurs tailles et leurs localisations varient très certainement selon les techniques employées pour les étapes de reverse transcription, de clonage et de séquençage.

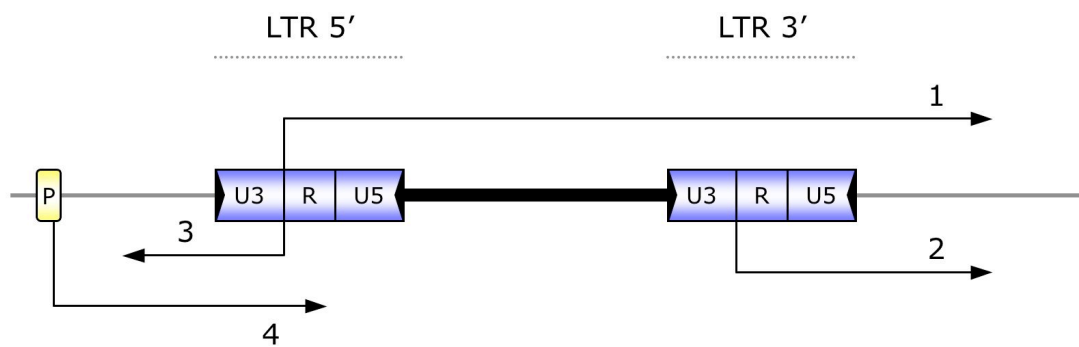
De plus, lorsqu'un rétrotransposon est faiblement exprimé, peu de séquences EST sont disponibles et elles apparaissent logiquement dans un nombre restreint de banques. Ceci augmente encore le biais observé dans la répartition de ces séquences sur l'élément. Par exemple, les séquences EST correspondant à l'élément *Crémant* sont toutes issues de la même banque et appartiennent toutes à la région interne, mais auraient peut-être été positionnées différemment si elles avaient été trouvées dans une autre banque. Le cas de l'élément *Gret1* est différent, puisqu'il s'agit d'un rétrotransposon de grande taille (> 10Kb) et qu'il est parmi les éléments les plus représentés dans les banques. Il faut tenir compte du fait que la synthèse d'un ADNc de plus de 10 kb est difficile et que les longs ARN messagers sont susceptibles d'être parfois partiellement dégradés. Il est donc possible que la position des EST des éléments de grande taille soit très fortement influencée par le mode de construction de la banque. Pour *Gret1*, plusieurs EST ont été localisées à l'extrémité 3' de la séquence et peuvent correspondre à des séquences obtenues en utilisant des amorces ciblant la queue poly(A) des transcrits, donc leur extrémité 3'.

Niveau de transcription variable selon l'ORF considéré

La majorité des séquences EST de *Gret1* correspondent à l'ORF2. Si le mode de production de la banque ne peut pas expliquer ce résultat, il faut alors envisager l'hypothèse selon laquelle cette ORF serait transcrite de manière indépendante et à un niveau plus élevé que le reste de la séquence, notamment l'ORF1. Ce mécanisme serait alors en contradiction avec l'hypothèse selon laquelle les rétrotransposons évolueraient vers des systèmes de transcription privilégiant la synthèse des protéines GAG par rapport aux protéines POL pour favoriser la formation de la VLP.

Transcrits chimériques

La présence d'EST correspondant à des transcrits chimériques peut expliquer la forte représentation des régions LTR dans les banques. Ces séquences correspondent à des fragments de LTR transcrits conjointement avec une séquence appartenant au génome hôte et peuvent avoir plusieurs origines (Figure 33). La production de tels transcrits a été déjà démontrée chez les plantes, notamment pour des rétrotransposons du blé (Kashkush *et al.*, 2003). La transcription d'ARN chimériques peut parfois influencer l'expression de gènes adjacents, modifier leur profil d'expression en augmentant leur niveau de transcription dans certains tissus ou en provoquant leur extinction par des mécanismes de silencing qui ciblent initialement les transcrits du rétroélément. Dans cette étude, la séquence des EST chimériques



1) La transcription du rétrotransposon est initiée entre dans la LTR 5' mais le site de terminaison est absent dans la LTR 3'. La synthèse de l'ARN se poursuit au delà de l'élément. **2)** La transcription est initiée dans la LTR 3', qui contient également un site d'initiation, et se poursuit au delà du site de terminaison. **3)** Un ARN antisens est transcrit à partir d'un promoteur *cryptique* présent dans la LTR 5'. **4)** Un promoteur de l'hôte initie la synthèse d'un transcrit qui recouvre une partie du rétrotransposon.

Figure 33 : Origines possibles des transcrits chimériques.

n'étaient cependant pas assez longues pour nous permettre de déterminer s'il s'agissait effectivement de la séquence de l'hôte, d'un gène en particulier, ou simplement de résidus non-éliminés de la séquence du vecteur ou de l'amorce utilisés pour le séquençage des EST.

I.5 - Conclusion

La structure et la diversité des 18 familles rétrotransposons sont aujourd'hui mieux connues que lorsque cette étude a été réalisée. Aussi, compte tenu des nouvelles données dont nous disposons, ces séquences EST pourront être analysées plus en détail pour déterminer les sites d'initiation de la transcription de chaque élément qui sont généralement localisés dans la région U5 ou R des LTR.

Finalement, les résultats de la recherche de séquences EST dans les banques de données doivent être pris avec précaution, mais ils donnent une première indication sur l'activité transcriptionnelle des rétrotransposons du génome de la vigne. Cette étude apporte notamment des informations sur le niveau de transcription des copies de plusieurs familles suite à un stress, un résultat qui pourra être confronté à ceux obtenus par RT-PCR.

Chapitre II : Évaluation du niveau de transcription par RT-PCR semi-quantitative

II.1 - Avant propos

Lorsque le travail de thèse a débuté, en janvier 2005, seuls trois rétrotransposons du génome de la vigne étaient décrits et peu d'informations étaient disponibles sur ces familles. Dans ces conditions, la mise au point des premières expériences de RT-PCR a été difficile puisque nous ne disposions d'aucune information sur le niveau d'expression de ces éléments, et qu'il était difficile de dessiner des amorces adaptées alors que presque aucune information sur la variabilité de ces familles n'était disponible, particulièrement dans le cas des familles *Gret1* et *Vine-1*, pour lesquelles une seule copie était identifiée. Par ailleurs, nous avons tenté d'amplifier par RT-PCR des transcrits de rétrotransposons de classe I en ciblant des domaines conservés de la région *pol*, mais les résultats obtenus à l'époque n'étaient pas répétables et le séquençage des produits RT-PCR ne confirmait pas l'appartenance de ces séquences à des éléments de classe I.

Comme nous l'avons vu précédemment, les premières données partielles du projet de séquençage du génome de la vigne n'ont été disponibles qu'au cours de l'année 2007, et il a fallu attendre le milieu d'année pour que les premiers contigs réellement exploitables nous permettent de caractériser de nouvelles familles de rétrotransposons. Au cours des deux premières années de thèse, des expériences de RT-PCR, réalisées sur les trois rétrotransposons *Gret1*, *Vine-1* et *Tvv1*, nous ont permis d'optimiser les différentes étapes du dispositif expérimental, c'est-à-dire la mise en place du matériel végétal, le conditionnement des plantes, le prélèvement des organes au cours du développement, les conditions d'application des stress, le protocole d'extraction des ARN totaux à partir de différents tissus et le protocole de RT-PCR puisque cette technique n'était pas utilisée au laboratoire. Finalement, lorsque les nouvelles familles de rétrotransposons ont été caractérisées au cours de l'année 2007, les échantillons végétaux étaient prêts et, avant la fin de la thèse, nous avons pu évaluer le niveau de transcription de ces éléments, à l'exception des familles *Kastel* et *Wintz* qui ont été caractérisées plus tard, et dont l'expression n'a donc pas pu être étudiée.

II.2 - Dispositif expérimental

Ce chapitre présente le dispositif expérimental mis en place (résumé dans la Figure 34) et les résultats obtenus par RT-PCR pour 16 familles de rétrotransposons du génome de la vigne.

II.2.1 - Le matériel végétal

a) *Choix de la variété Riesling*

Le choix du Riesling comme matériel d'étude s'est imposé à nous pour plusieurs raisons. Tout d'abord, la famille *Tvv1* a été découverte par une stratégie de marche sur chromosome réalisée sur ce génome, nous donnant accès à la séquence partielle de nombreuses copies (Pelsy and Merdinoglu, 2002). Ces données nous ont permis de dessiner des amorces consensus permettant d'amplifier par PCR le maximum d'éléments de cette famille. Changer de variété de vigne serait donc revenu à prendre le risque de diminuer le nombre de copies amplifiées par ces amorces, sachant que la séquence d'une copie est susceptible d'avoir accumulé des mutations depuis son insertion dans le génome et donc d'avoir divergé de son homologue présente dans une autre variété.

Ensuite, le séquençage du génome de l'accession PN40024 n'étant à l'époque qu'un projet, nous n'étions pas certains de disposer de séquences génomiques avant la fin de la thèse. De plus, ce génome est susceptible d'avoir perdu environ 50% de ses copies d'éléments transposables lors des cycles d'autofécondation qui ont été réalisés pour le rendre homozygote à 93%. Donc, ce génome a probablement perdu des copies actives, ce qui aurait rendu la détection de transcrits de rétrotransposons plus difficile.

Enfin, la multiplication de l'accession PN40024 a longtemps été exclusivement réservée au projet de séquençage. Dans ces conditions, la mise en place d'un grand nombre de boutures était plus facile à réaliser à partir de bois de Riesling, dont l'INRA dispose en plus grande quantité.

b) *Mise en place de collections de boutures*

Des boutures de la variété Riesling, clone agréé n°49, ont donc été réalisées. Pour cela, 150 sarments aoûtés ont été prélevés au mois de janvier sur des plants de vignes adultes. Les bois ont été enveloppés dans un sac plastique, puis conservés à 4°C pour les protéger de

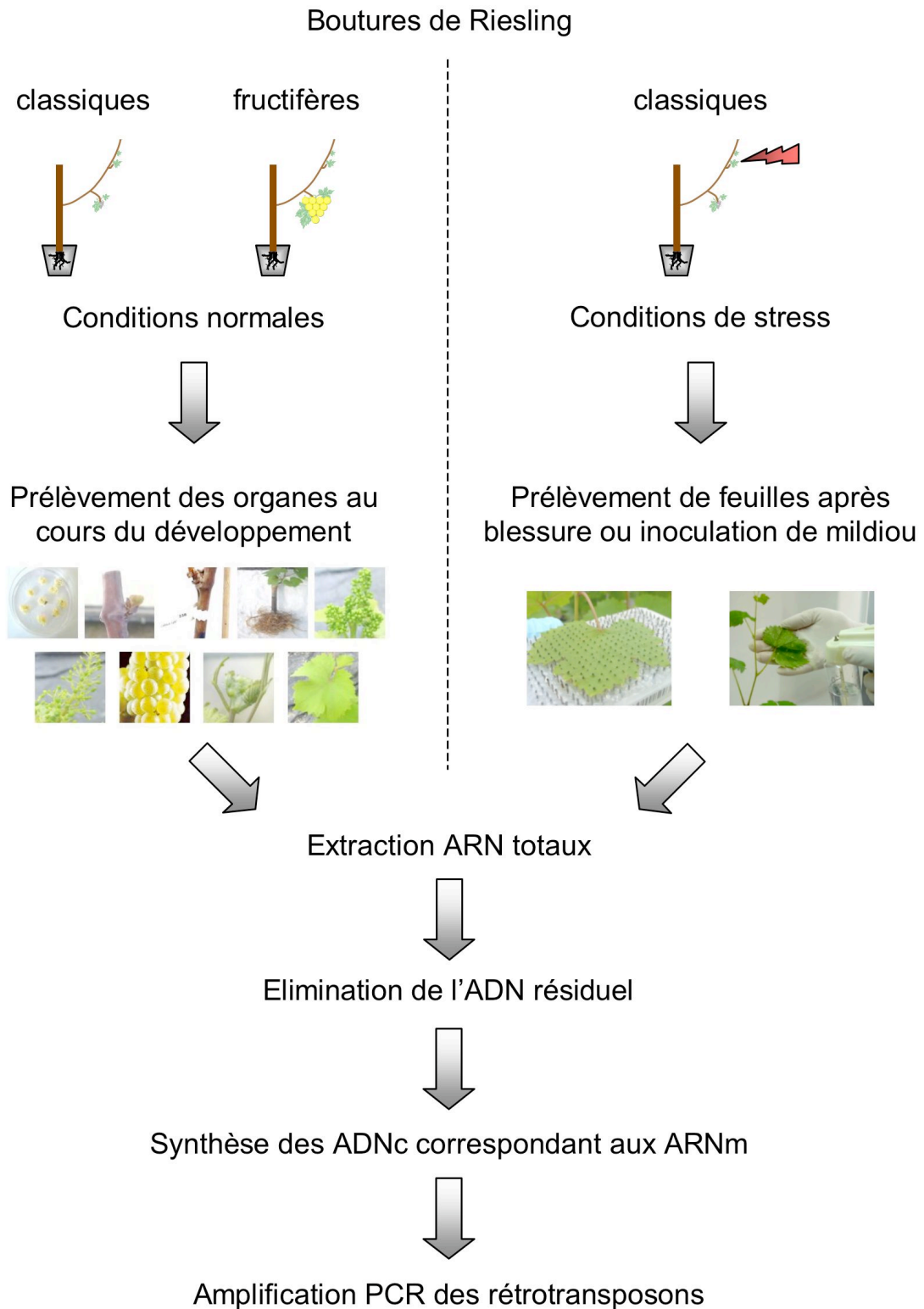


Figure 34 : Schéma récapitulatif du dispositif expérimental mis en place pour évaluer la transcription des rétrotransposons par RT-PCR semi-quantitative.

l'assèchement. Au mois de février, les rameaux ont été débités pour obtenir 500 boutures qui ont été conservées au froid pour respecter le délai de dormance des bourgeons. Au mois d'avril, les boutures ont été mises en pot, puis placées en serre sur une nappe chauffante pour favoriser le développement des racines (Figure 35). Celles-ci se sont développées en premier car la croissance du bourgeon a été naturellement ralentie par une température ambiante plus basse en serre à cette période de l'année. Le débourrement des premiers bourgeons s'est produit environ cinq à sept semaines après la mise en pot.

Boutures « classiques »

Pour obtenir tous les organes végétatifs nécessaires à l'étude, deux tiers des boutures ont été conduites de manière « classique », c'est-à-dire sans intervention particulière (le protocole est présenté dans la Partie V : Matériels et méthodes). Ces plantes ont permis des prélèvements des méristèmes de racines, de bois, de bourgeons juste avant leur débourrement (ou bourgeons « dans le coton »), de méristèmes apicaux et de feuilles (Figure 36).

Boutures « fructifères »

Pour obtenir les organes reproducteurs dès la première année de développement, le tiers des boutures restant a été conduit en boutures « fructifères », c'est-à-dire en réalisant des opérations visant à modifier les équilibres nutritifs des différents organes susceptibles d'entrer en croissance. En éliminant, dès leur apparition, les premières feuilles et l'apex du rameau, les réserves énergétiques des boutures ont été redirigées pour privilégier la première inflorescence. L'entre-cœur apparu en face de l'inflorescence a été conservé. Le rameau et les feuilles qui se sont développés ont servi à alimenter l'inflorescence en produits photosynthétiques. Les boutures ont ensuite été conduites normalement et l'inflorescence a fleuri environ six semaines après le débourrement. Ces boutures « fructifères » ont permis d'obtenir successivement les organes suivants : des boutons floraux, des baies vertes à la période de la nouaison, et des pellicules de baies mûres (Figure 36).

c) Prélèvement des échantillons d'organes

Pour prévenir l'apparition de maladies, toutes les plantes en serre ont été traitées plusieurs fois au cours de leur développement avec des produits phytosanitaires classiques. Cependant, tous les prélèvements ont été réalisés au minimum 20 jours après un traitement afin d'éviter toute interférence, et entre 7h et 9h du matin pour que les plantes soient dans un état



Figure 35 : Installation des boutures sur nappe chauffante en serre.

physiologique relativement homogène, même si tous les organes n'étaient pas prélevés le même jour. Les organes ainsi que les plantes sur lesquelles les prélèvements ont été effectués étaient tous sains, et présentaient le même niveau de développement. Par exemple, les feuilles étaient en position n°4 ou 5 en partant de l'apex, et leurs tailles étaient toutes similaires. Immédiatement après leur prélèvement, les échantillons d'organes ont été plongés dans de l'azote liquide, puis conservés à -80°C.

Compte tenu du cycle de développement de la vigne, tous les organes ne sont pas disponibles en même temps. Certains organes, comme les racines et le bois, ont été prélevés sur la même bouture. Les méristèmes apicaux et les feuilles ont été prélevés simultanément sur d'autres boutures. Les boutons floraux et les baies ont ensuite été prélevés sur différentes boutures fructifères, lorsque celles-ci avaient atteint le stade de développement attendu. Dans tous les cas, les plantes sur lesquelles un ou plusieurs organes avaient été prélevés, ont été systématiquement éliminées du lot étudié.

d) Embryons somatiques

Pour compléter cette gamme d'organes de vigne, des embryons somatiques nous ont été donnés par l'équipe du Laboratoire de Biologie du Développement de la Vigne de l'INRA de Colmar (Figure 36). L'embryogenèse somatique est un procédé utilisé chez la vigne pour la régénération de plantes à partir d'un type de tissus particulier. Ces embryons, ou cals embryogènes, étaient des cals stabilisés issus d'embryogenèse somatique obtenus à partir de culture de filet d'anthers de Riesling (Perrin *et al.*, 2004).

II.2.2 - Stress biotique et abiotique

a) Préparation des plantes

L'évolution du niveau de transcription des rétrotransposons en réponse à l'induction par des stress a été évaluée sur des feuilles soumises à deux types de stress potentiellement subis par la vigne lors de sa culture. Pour cela, des boutures « classiques », dont le développement était uniforme et dont les feuilles étaient parfaitement saines, ont été placées dans un caisson climatisé maintenu à une température de 23°C, avec une photopériode de 16h de lumière pour 8h d'obscurité. La température extérieure relevée au moment du transfert des plantes en

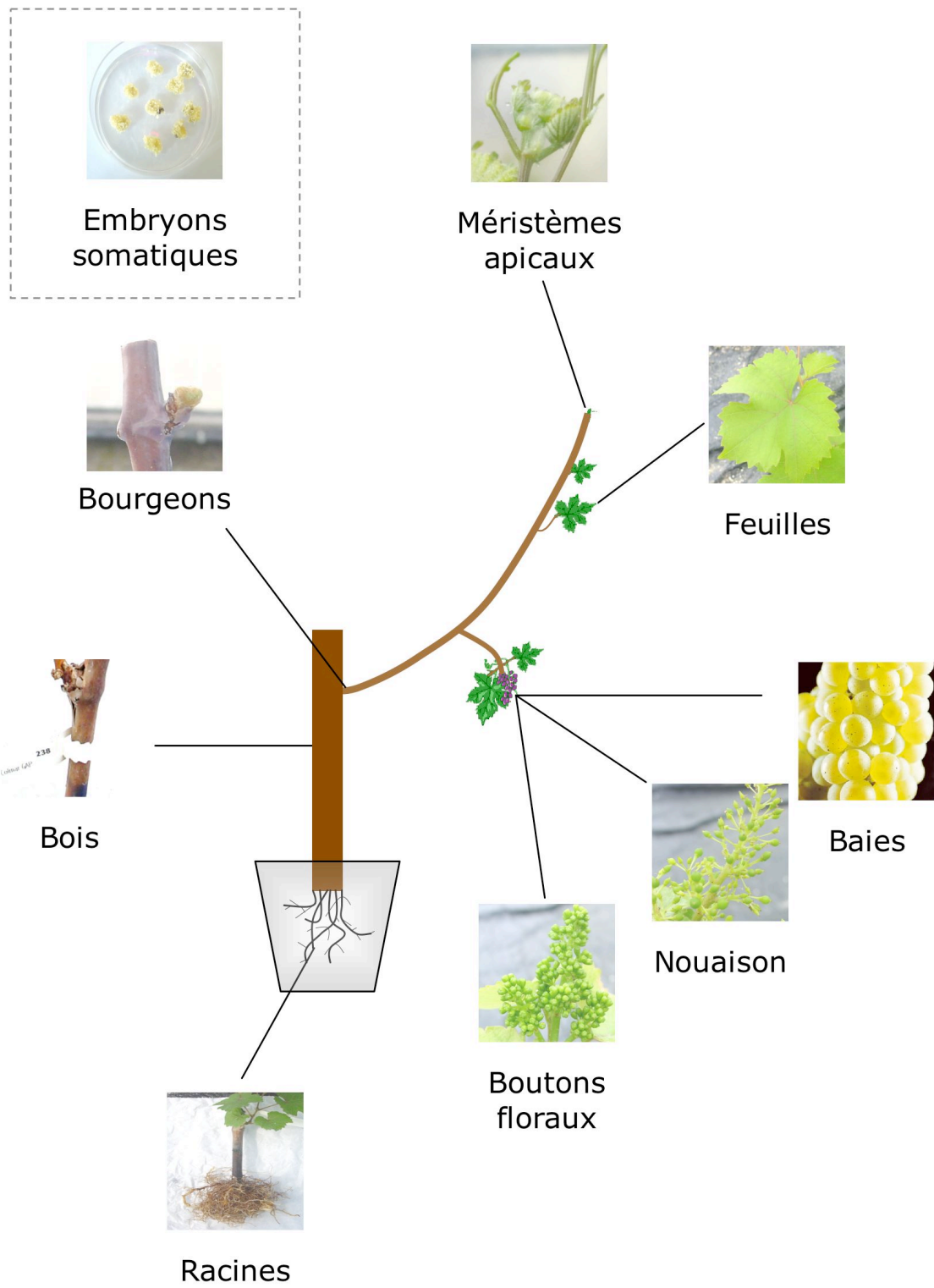


Figure 36 : Organes prélevés sur les boutures et embryons somatiques de Riesling.

caisson climatique avoisinait les 20°C. Les plantes n'ont donc *a priori* pas subi de choc thermique important lors de leur transfert. De plus, les boutures ont été laissées quatre jours en caisson avant l'application des stress, afin de minimiser l'influence du passage des plantes en caisson climatisé sur les résultats de l'expérience.

b) Application des stress

Deux stress, abiotique et biotique, ont ensuite été appliqués sur les feuilles et chaque expérience de stress a été réalisée deux fois de façon totalement indépendante.

Blessure mécanique

Le stress abiotique, la blessure mécanique, avait déjà été utilisé pour étudier l'induction de l'expression de rétrotransposons chez les plantes : pour étudier l'expression de *Tto1*, des morceaux de feuilles de tabac ont été découpés puis coupés à 20 reprises à l'aide d'une pince (Takeda *et al.*, 1998) ; dans le cas de l'élément *Ogre*, la surface de feuilles de pois a été écorchée à l'aide de papier de verre (Neumann *et al.*, 2003) ; pour l'étude de *CIRE1*, des feuilles d'oranger ont été blessées par des coupures réalisées au scalpel (Rico-Cabanas and Martinez-Izquierdo, 2007). Pour éviter les biais pouvant être occasionnés par le prélèvement de la feuille et le découpage de disques foliaires, nous avons choisi d'appliquer la blessure sur des feuilles entières non détachées. Pour permettre la répétition du même stress sur plusieurs feuilles, la blessure a été réalisée à l'aide d'une plaque munie de pointes qui, lorsqu'elle a été appliquée contre les feuilles, a provoqué leur perforation à raison d'environ 15 blessures par centimètre carré (Figure 37).

Des feuilles saines et non-blessées ont été prélevées au début de l'expérience pour servir de témoins. Des feuilles ont ensuite été prélevées 6h, 12h, 24h et 48 heures après leur blessure.

Attaque par un pathogène

Le stress biotique a consisté à infecter la feuille par le mildiou de la vigne *Plasmopara viticola*, un pathogène biotrophe. Dans ce cas, les feuilles ont été inoculées par pulvérisation d'une solution de sporanges de mildiou sur leur face inférieure (Figure 38). Cette technique est utilisée au laboratoire pour multiplier le mildiou et réaliser certains tests de phénotypage (Diez-Navajas *et al.*, 2006). Les feuilles infectées ont ensuite été enveloppées dans un sachet plastique pour conserver une humidité favorable au développement du mildiou (Figure 39).

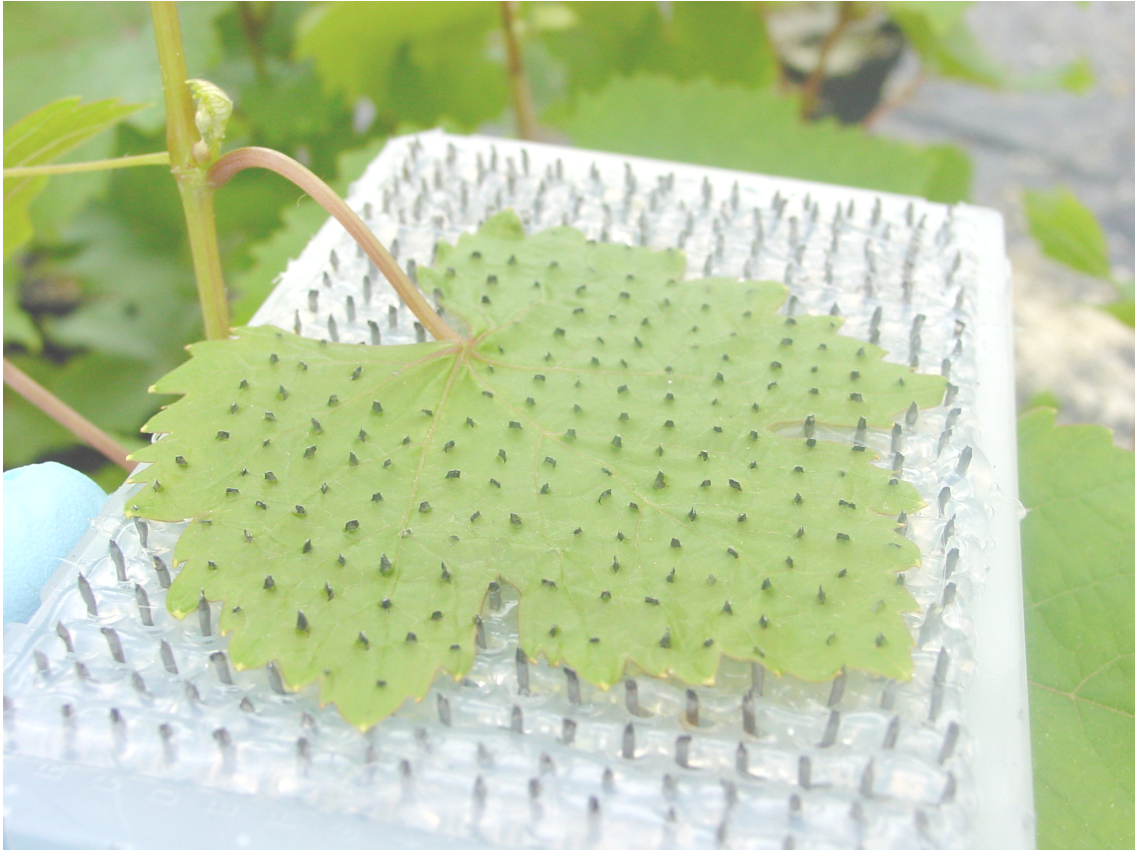


Figure 37 : Blessure mécanique d'une feuille par application d'une plaque munie de pointes.



Figure 38 : Inoculation d'une feuille par pulvérisation d'une solution de sporanges de mildiou (*P. viticola*) sur sa face inférieure.



Figure 39 : Les feuilles inoculées par la solution de mildiou sont recouvertes d'un sachet en plastique pour préserver une humidité suffisante au développement du pathogène.

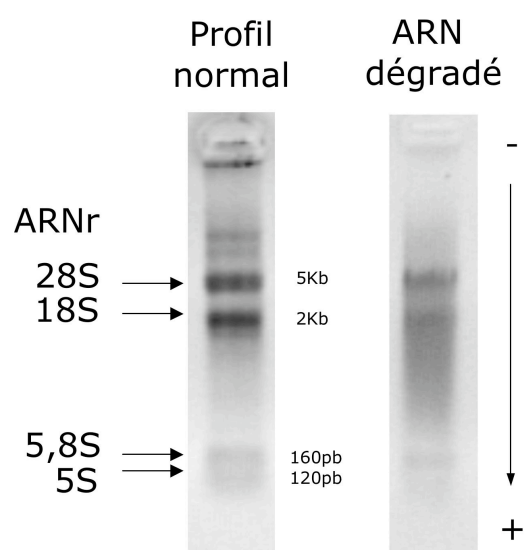


Figure 40 : Exemples de profils obtenus après migration des ARN totaux sur gel d'agarose.

La température du caisson climatisé (23°C) a été choisie pour permettre le développement optimal du mildiou (Diez-Navajas *et al.*, 2006).

Des feuilles non-inoculées mais enveloppées dans les sachets plastiques ont été prélevés 12h après le début de l'expérience pour servir de témoins sains, alors que des feuilles infectées ont été prélevées 6h, 12h, 24h et 48 heures après l'inoculation.

Témoin d'infection

En parallèle de l'inoculation des feuilles de Riesling, des feuilles de Muscat ottonel, une variété très sensible au mildiou, ont été inoculées. Ces feuilles ont été observées 7 jours après l'inoculation pour vérifier *a posteriori* que la solution d'inoculum était suffisamment concentrée, que la souche de mildiou était agressive, et que le pathogène s'était bien développé sur les feuilles.

II.2.3 - Extraction des ARN totaux

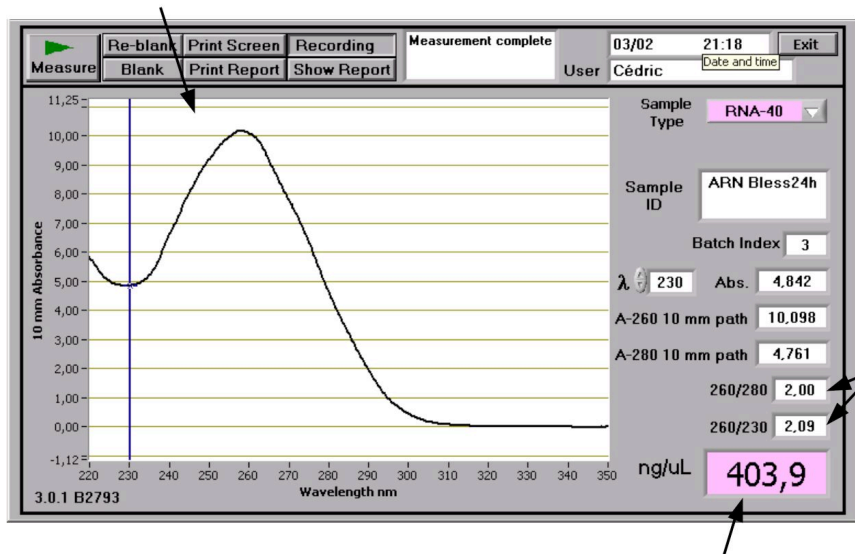
a) Extraction

Les extractions d'ARN totaux ont été réalisées sur des mélanges de 4 à 6 échantillons selon l'organe, représentant au total 100 mg de matière végétale, en utilisant le kit RNeasy Plant Mini Kit (QUIAGEN). Plusieurs étapes du protocole ont été modifiées par rapport aux recommandations du fournisseur afin d'optimiser le rendement et la qualité des extractions (voir Partie V : Matériels et méthodes). L'ARN total correspond à un mélange de tous les ARN présents au niveau cellulaire, c'est-à-dire les ARN ribosomiaux (ARNr), les ARN messagers (ARNm) et les ARN de transfert (ARNt). Les ARNm ne représentent généralement que 3 à 5 % du total.

b) Elimination de l'ADN génomique résiduel

Une étape d'hydrolyse enzymatique de l'ADN a été réalisée en utilisant le kit RNase-Free DNase Set (QUIAGEN) pour s'assurer de l'absence de toute trace d'ADN pouvant perturber les résultats de l'étude.

Courbe d'absorbance de l'échantillon à différentes longueurs d'ondes



Rapports d'absorbances donnant une indication sur la proportion de résidus phénoliques et de protéines résiduels

Concentration de l'échantillon

Figure 41 : Test de la qualité des ARN totaux extraits (courbe de densité optique et rapports d'absorbance d'un échantillon d'ARN extrait à partir de feuilles blessées).

c) Vérification de la qualité des ARN extraits

L'intégrité des ARN totaux a été évaluée par électrophorèse sur gel d'agarose : le profil attendu pour des ARN totaux de bonne qualité présente des bandes d'ARN ribosomiques intactes, aucun smear témoignant de dégradations, et aucune bande de haut poids moléculaire pouvant correspondre à de l'ADN génomique résiduel (Figure 40). Lorsque cela était nécessaire, de nouvelles extractions ont été réalisées jusqu'à obtenir des échantillons d'ARN sans dégradation.

La qualité des échantillons d'ARN a été également évaluée en mesurant leur courbe d'absorbance à différentes longueurs d'ondes. Seuls les échantillons d'ARN présentant une courbe densité optique correcte et des rapports d'absorbance 260/230 et 260/280 compris entre 1,80 et 2,10 ont été conservés (Figure 41). Ces vérifications systématiques étaient indispensables pour s'assurer de l'absence de composés phénoliques ou de protéines indésirables susceptibles de perturber les étapes de RT-PCR ultérieures.

II.2.4 - Synthèse des ADN complémentaires (ADNc)

a) Etalonnage des quantités d'ARN

En partant du même poids de matière fraîche, il est souvent difficile d'obtenir les mêmes rendements d'extraction à partir de tissus différents. En effet, les rendements d'extraction obtenus à partir de bois ou de pellicules de baies sont, par exemple, nettement inférieurs à ceux obtenus à partir d'organes en croissance comme les bourgeons ou les feuilles. Nous avons donc décidé d'utiliser des quantités d'ARN différentes selon l'étude considérée.

Ainsi, 2 μ g d'ARN total par échantillon ont été utilisés pour synthétiser les ADNc qui ont servi à l'évaluation du niveau de transcription des rétrotransposons dans les organes en conditions normales de croissance.

Par contre, les extractions d'ARN étant plus efficaces et les rendements plus constants à partir de feuilles, nous avons utilisé 4 μ g d'ARN total de feuilles (saines, blessées ou infectées) dans le cadre de l'étude de l'influence des stress.

b) Synthèse des ADNc simple-brin

La synthèse du 1^{er} brin des ADN complémentaires (ou ADNc) est suffisante pour effectuer la quantification des transcrits par PCR. Des ADNc simple-brin ont donc été synthétisés avec le kit Revert AidTM - First Strand cDNA Synthesis Kit (Fermentas Life Science) en utilisant des amorces oligo(dT) pour synthétiser spécifiquement des ADNc correspondant aux ARN messagers poly(A). L'amorçage par des amorces oligo(dT) est le seul qui permette d'obtenir des ADNc pleine longueur. Comme précédemment, l'intégrité et la qualité des préparations d'ADNc ont été testées grâce au spectrophotomètre.

II.2.5 - Témoin interne : le gène de l'actine

Plusieurs problèmes peuvent apparaître lors de l'évaluation du niveau d'expression d'un transcrit par RT-PCR semi quantitative. Tout d'abord, l'ARN est une molécule relativement fragile et l'évaluation de la quantité d'un transcrit ne tient pas compte de son niveau de dégradation, d'autant plus que les ARNm ne représentent qu'une faible proportion des ARN totaux. Ensuite, il est difficile de corrélérer une quantité d'ARN à un nombre exact de cellules effectivement étudiées.

Pour ces raisons, l'utilisation d'un transcrit endogène est indispensable. Ce transcrit doit refléter la quantité d'ARN analysée, sa qualité, et l'efficacité de l'étape de transcription inverse. Son expression doit être constante d'une plante et d'un type cellulaire à l'autre, et ne doit pas être affectée par l'état physiologique de la plante. Son taux de transcription et son rendement d'amplification doivent être comparables à celui des cibles. Pour s'approcher au plus près de ces conditions idéales, le gène de l'actine a été utilisé comme référence car il est considéré comme étant constitutivement et uniformément exprimé dans tous les organes.

a) Test PCR sur ARN totaux

L'absence de bande de haut poids moléculaire après migration d'un échantillon d'ARN sur gel d'agarose ne prouvant pas de façon certaine que tout l'ADN génomique a été correctement éliminé de l'échantillon, un test d'amplification du gène de l'actine a été réalisé par PCR, directement sur les ARN totaux. Les amorces utilisées pour la PCR ne peuvent pas s'hybrider correctement sur une séquence ARN et aucune amplification ne doit être obtenue sur les échantillons ARN non contaminés. Lorsqu'une amplification était obtenue, prouvant la

persistance de traces ADN dans l'échantillon d'ARN totaux, celui-ci était traité avec le kit Turbo DNase-free™ (AMBION) et repassait ensuite les tests d'électrophorèse et de PCR actine.

b) Evaluation de l'intégrité des ADNc

Les amorces utilisées pour amplifier l'actine ayant été dessinées de manière à encadrer un intron, les amplifications PCR issues d'une matrice ADN d'une part, et d'une matrice ADNc d'autre part, se distinguent par une différence de taille des fragments amplifiés, les transcrits ayant normalement subi une phase d'épissage des introns. Les fragments amplifiés à partir d'ADNg ont une taille de 515 pb, alors que les fragments issus d'ADNc étaient de 426 pb. L'amplification du gène de l'actine, qui a été confirmée par le clonage et le séquençage des produits obtenus par PCR et par RT-PCR (Figure 42), a donné comme prévu des fragments de tailles différentes selon la matrice utilisée (Figure 43).

II.2.6 - Evaluation de la quantité relative de transcrits par PCR semi-quantitative

a) Dessin des amorces spécifiques des rétrotransposons

Des amorces spécifiques de 16 familles de rétrotransposons du génome de la vigne ont été dessinées à partir des séquences génomiques disponibles et des EST précédemment identifiées. Pour amplifier spécifiquement chaque famille de rétrotransposon, nous avons choisi de dessiner des couples d'amorces dans la région *gag* en dehors des motifs nucléiques les plus conservés entre les familles. Il a été démontré que cette région évolue jusqu'à deux fois plus rapidement que la région RT (McClure *et al.*, 1988) et qu'elle permet de différencier les familles de rétrotransposons mais reste suffisamment conservée entre les copies appartenant à une même famille. Lorsque la région *gag* n'était pas délimitée, comme c'était le cas pour les éléments *V*, les couples d'amorces ont été dessinés dans la région 5' de la séquence interne du rétrotransposon.

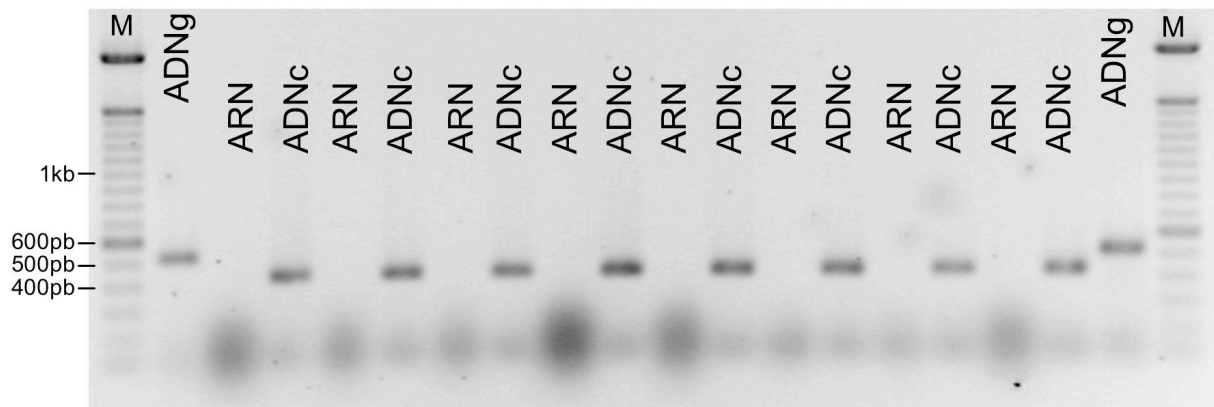
Tous les couples d'amorces dessinés pour cibler spécifiquement les rétrotransposons ont permis l'amplification de fragments aux tailles attendues à partir d'ADN génomique de Riesling.

		10 20		30	40	50		
I1ACT-ADNc	5	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
I1ACT-ADNc	6	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
I1ACT-ADNc	7	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
I1ACT-ADNc	8	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
I2ACT-ADNg	2	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
I2ACT-ADNg	3	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
I2ACT-ADNg	4	1	TACAATGAGC	TTCGGGTTGC	CCCTGAAGAA	CACCCAGTGC	TTCTAACTGA	50
			60	70	80	90	100	
I1ACT-ADNc	5	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
I1ACT-ADNc	6	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
I1ACT-ADNc	7	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
I1ACT-ADNc	8	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
I2ACT-ADNg	2	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
I2ACT-ADNg	3	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
I2ACT-ADNg	4	51	GGCACCCCTC	AACCCCAAGG	CCAACAGAGA	AAAGATGACT	CAAATCATGT	100
			110	120	130	140	150	
I1ACT-ADNc	5	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
I1ACT-ADNc	6	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
I1ACT-ADNc	7	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
I1ACT-ADNc	8	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
I2ACT-ADNg	2	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
I2ACT-ADNg	3	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
I2ACT-ADNg	4	101	TTGAGACTTT	TAATGTGCCT	GCCATGTATG	TTGCCATTCA	GGCTGTCTT	150
			160	170	180	190	200	
I1ACT-ADNc	5	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	-----	-----	200
I1ACT-ADNc	6	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	-----	-----	200
I1ACT-ADNc	7	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	-----	-----	200
I1ACT-ADNc	8	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	-----	-----	200
I2ACT-ADNg	2	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	TCGTATTAAA	ATCTTTTTGG	200
I2ACT-ADNg	3	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	TCGTATTAAA	ATCTTTTTGG	200
I2ACT-ADNg	4	151	TCTCTATATG	CCAGTGGGCG	TACAACTGGT	TCGTATTAAA	ATCTTTTTGG	200
			210	220	230	240	250	
I1ACT-ADNc	5	201	-----	-----	-----	-----	-----	250
I1ACT-ADNc	6	201	-----	-----	-----	-----	-----	250
I1ACT-ADNc	7	201	-----	-----	-----	-----	-----	250
I1ACT-ADNc	8	201	-----	-----	-----	-----	-----	250
I2ACT-ADNg	2	201	ATTGGAATTT	TCATTTTCTC	TTTCTGTTGT	TAGTGTAGCC	TGAGTTCTAA	250
I2ACT-ADNg	3	201	ATTGGAATTT	TCATTTTCTC	TTTCTGTTGT	TAGTGTAGCC	TGAGTTCTAA	250
I2ACT-ADNg	4	201	ATTGGAATTT	TCATTTTCTC	TTTCTGTTGT	TAGTGTAGCC	TGAGTTCTAA	250
			260	270	280	290	300	
I1ACT-ADNc	5	251	-----	-----A	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
I1ACT-ADNc	6	251	-----	-----A	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
I1ACT-ADNc	7	251	-----	-----A	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
I1ACT-ADNc	8	251	-----	-----A	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
I2ACT-ADNg	2	251	CTTGCTCCTT	TGACTAGGTA	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
I2ACT-ADNg	3	251	CTTGCTCCTT	TGACTAGGTA	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
I2ACT-ADNg	4	251	CTTGCTCCTT	TGACTAGGTA	TTGTGCTGGA	TTCTGGTGAT	GGTGTGAGTC	300
			310	320	330	340	350	
I1ACT-ADNc	5	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350
I1ACT-ADNc	6	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350
I1ACT-ADNc	7	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350
I1ACT-ADNc	8	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350
I2ACT-ADNg	2	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350
I2ACT-ADNg	3	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350
I2ACT-ADNg	4	301	ACACTGTGCC	AATTTATGAA	GGTTATGCAC	TTCCCCATGC	TATCCTTCGT	350

		360	370	380	390	400		
I1ACT-ADNc	5	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
I1ACT-ADNc	6	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
I1ACT-ADNc	7	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
I1ACT-ADNc	8	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
I2ACT-ADNg	2	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
I2ACT-ADNg	3	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
I2ACT-ADNg	4	351	CTTGACCTTG	CTGGACGTGA	CCTCACTGAT	GCTCTGATGA	AGATTCTAAC	400
			410	420	430	440	450	
I1ACT-ADNc	5	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
I1ACT-ADNc	6	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
I1ACT-ADNc	7	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
I1ACT-ADNc	8	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
I2ACT-ADNg	2	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
I2ACT-ADNg	3	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
I2ACT-ADNg	4	401	TGAGAGGGGT	TACATGTTCA	CCACTACTGC	TGAACGGGAA	ATTGTCCGGG	450
			460	470	480	490	500	
I1ACT-ADNc	5	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
I1ACT-ADNc	6	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
I1ACT-ADNc	7	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
I1ACT-ADNc	8	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
I2ACT-ADNg	2	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
I2ACT-ADNg	3	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
I2ACT-ADNg	4	451	ACATGAAGGA	AAAGCTTGCA	TATGTTGCTC	TTGACTACGA	GCAAGAGCTG	500
			510					
I1ACT-ADNc	5	501	GAAACTGCAA	AGAGC	550
I1ACT-ADNc	6	501	GAAACTGCAA	AGAGC	550
I1ACT-ADNc	7	501	GAAACTGCAA	AGAGC	550
I1ACT-ADNc	8	501	GAAACTGCAA	AGAGC	550
I2ACT-ADNg	2	501	GAAACTGCAA	AGAGC	550
I2ACT-ADNg	3	501	GAAACTGCAA	AGAGC	550
I2ACT-ADNg	4	501	GAAACTGCAA	AGAGC	550

Les séquences I1ACT-ADNc (426pb) et I2ACT-ADNg (515pb) correspondent respectivement à des séquences obtenues à partir d'ADN complémentaire et à partir d'ADN génomique de Riesling. La position des amorces actine sens et antisens est indiquée par des cadres noirs.

Figure 42 : Alignement des séquences actine obtenues après le clonage de fragments amplifiés à partir de matrices ADNc ou ADNg.



M : marqueur de taille

Figure 43 : Exemples d'amplifications PCR du gène de l'actine sur des échantillons d'ADN génomique, d'ARN totaux et d'ADN complémentaires.

Deux couples d'amorces supplémentaires ont été dessinés pour amplifier la région UTL des éléments *Tvv1*. Cette région est considérée comme très variable au sein d'une famille (Pelsy, 2007) et le séquençage de l'UTL des transcrits doit nous permettre de différencier les copies ou un petit nombre de copies transcrites au sein de cette famille.

Les amorces ont été définies avec le logiciel Primer3, sur le site : http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi. Le Tableau 10 présente les principaux critères utilisés pour choisir les couples d'amorces. Leurs séquences et leurs positions sur les éléments ciblés sont données dans le Tableau 11.

b) Gènes témoins de l'effet des stress

Pour confirmer l'effet des stress sur la physiologie de la cellule, deux couples d'amorces supplémentaires, nommés *PR1* et *Def1* ont été dessinés pour amplifier des familles de gènes PR (pour Pathogen Related). Les niveaux de transcription de ces gènes varient en réponse à des changements de conditions physiologiques, notamment lors de l'attaque de pathogènes et en réponse à des stress abiotiques comme la blessure (Van Loon and Van Strien, 1995; Roetschi *et al.*, 2001). L'évaluation de la quantité de transcrits ciblés par ces amorces permet donc d'estimer le niveau de transcription attendu pour des gènes sur-exprimés après un stress. Pour définir ces amorces, les mots clefs « PR1 » et « defensine » ont été utilisés pour rechercher des séquences de vigne dans la banque de données TIGR (Pere Mestre, communication personnelle). Les séquences nucléiques obtenues pour chaque mot clef ont été traduites et puis alignées par ClustalW. Des amorces ont été dessinées dans les domaines conservés pour permettre l'amplification des transcrits de ces deux familles de gènes (Tableau 11).

c) Etalonnage des quantités d'ADNc

En règle générale, les résultats obtenus par PCR semi-quantitative ne tiennent pas compte des rendements d'amplification particuliers obtenus pour chaque couple d'amorces. L'efficacité de détection d'un rétrotransposon est pourtant influencée par des facteurs tels que la séquence nucléotidique des amorces, les températures d'hybridation et la séquence et la taille de l'amplicon. Ainsi, les signaux obtenus dans différentes conditions (organes ou feuilles stressées) pour un même couple d'amorces, donc pour un rétrotransposon donné, peuvent être

Position de l'amplicon sur la séquence de référence	Domaine <i>gag</i> ou à proximité
Taille de l'amplicon	150 à 550 pb
Température T _m des amorces	55 à 63°C
Taille des amorces	21 ou 22 pb
Complémentarité des amorces	< 8 bases complémentaires
Complémentarité des amorces en 3'	< 3 bases complémentaires
Composition des amorces	41 < %GC < 70

Tableau 10 : Critères utilisés pour le choix des amorces de rétrotransposons.

Famille	Taille de l'amplicon (pb)	Position sur l'élément	Séquences des amorces	
			sens	antisens
<i>Brand</i>	209	<i>gag</i>	TTCTGCCACATMTARGGTTGTG	ACCWYTTGCCACATACRTACC
<i>Cremant</i>	216	<i>gag</i>	GGCTACRTAAATGGGAGACAGC	GGTCCTGTATCACACAAAACGA
<i>Edel</i>	261	<i>gag</i>	CACTCCACACCGAGAAATCATA	CCCTAGCTCCCTTTGTAACCTT
<i>Gans</i>	282	<i>gag</i>	CTGAAGGTGATTCGGTACAACA	AACGAGCTCCACAAGAAGAGAC
<i>Gentil</i>	200	<i>gag</i>	GAGAAAGCCTCTGCATTGGTAG	ATGACTGGGGTTGTGTTACCAT
<i>Gret1</i>	313	<i>gag</i>	GCTGCTATGCAAAGTATTTCCC	AACCGCTTCACAAATTCCTTA
<i>Huben</i>	321	<i>gag</i>	GAAGTTGGGCCACACTAAAGAC	GACCWGAGGCTTGAAATTAGA
<i>Noble</i>	259	<i>gag</i>	GGTTTCAGGAAGTGGAAGATTG	AGTCAAGTTGCTGCCAATACCT
<i>Rangen</i>	173	<i>gag</i>	CACATGATCCTCACATGGATTG	GGTTCTTGCYTGAGTTGATGYA
<i>Tvv1</i>	526	<i>gag</i>	GCTGATATAGTGCCTATTGTGT	CAGAAAGAATCTGAGATTTTGC
<i>Vine1</i>	301	<i>gag</i>	AGGTGATGAACTTGATTGCGAGA	GCTCCTTCTGCAGTATCTCCTT
<i>V1</i>	260	début ORF	TGGCATATGGAGYGATACTTCG	TAGTCGCRATCGAWWCTGTGT
<i>V11</i>	293	début ORF	AAGAGARCCTACCCACAAAG	CACTCATCTTCATGATKTGMTC
<i>V12</i>	220	début ORF	GAGGGCTACCATAGTCACCTTG	RCTTCTTTMTGGCAAGAACCTG
<i>V13</i>	255	début ORF	CTCGAGAGGTGTGGAGTTCTCT	TGGCAAAGGACWGAGAGGTTAT
<i>V14</i>	179	milieu ORF	RGAGGCATGATARGTTCTTTGC	TCACYCTTAGGATAYCATAGGC
<i>Tw1 (UTL)</i>	variable	UTL	ATTACCTTGATTGCGTTACATG	ACAGAATTCTTACTCTCTCCA
<i>Actine</i>	426* ou 515**	-	TACAATGAGCTTCGGGTTGC	GCTCTTTCAGTTTCCAGCT
<i>PR1</i>	360	-	CAGGACTACCTCAATGCTCAC	TAGTTGCATGTGACGAACCAC
<i>Def1</i>	190	-	TTCTCCTCTTCATGACCACTG	AACAATGCTTAGTGCAGAAGC

* sur ADNc ** sur ADNg

Tableau 11 : Tailles et positions des couples d'amorces sur la séquence du rétrotransposon.

comparés avec une précision relativement bonne. A l'inverse, il est difficile de comparer des signaux obtenus par deux couples d'amorces différents dans la même condition, sinon en supposant que l'efficacité des rendements d'amplification PCR sont équivalents entre les deux couples.

En supposant que les étapes de transcription inverse et d'amplification PCR ont des rendements similaires quel que soit l'échantillon, l'intensité du signal détecté par RT-PCR pour l'actine doit refléter la quantité d'ARN de départ. Ainsi, afin d'évaluer les quantités relatives de transcrits détectées pour les rétrotransposons en les comparant à celles détectées pour le gène de l'actine, les signaux obtenus lors de l'amplification du gène de l'actine ont été étalonnés : après 24 cycles PCR, les échantillons ARN présentant un signal actine plus fort que les autres ont été dilués pour obtenir un signal non-saturant et homogène chez tous les échantillons de la même série. Cette précaution prise, la quantité d'ADNc utilisée pour estimer le niveau de transcription des rétrotransposons a été considérée comme équivalente dans tous les échantillons d'une même série. Ensuite, la quantité relative des transcrits de chaque famille de rétrotransposons a été évaluée en réalisant 26 cycles PCR.

II.3 - Résultats

II.3.1 - Détermination du niveau de transcription dans les organes

Le signal obtenu pour l'amplification des transcrits du gène de l'actine était légèrement plus fort pour deux organes, les méristèmes apicaux et les feuilles, indiquant que la quantité d'ADNc de ces échantillons était probablement un peu plus importante que pour les autres organes de la série (Figure 44). Pour les autres organes, les amplifications étaient homogènes, prouvant que les quantités d'ADNc de départ étaient très proches.

Comme le montrent les résultats présentés dans la Figure 44, les amplifications réalisées à partir des ADNc étalonnés par l'actine donnent des résultats très différents selon les familles de rétrotransposons considérées. Tout d'abord, quelque soit l'organe considéré, aucune amplification n'a été obtenue pour les éléments *Brand*, *Crémant*, *Rangen*, *V11*, *V12*, *V13* et *V14*. A l'inverse, les ADNc préparés à partir d'extractions d'ARN d'embryons somatiques ont

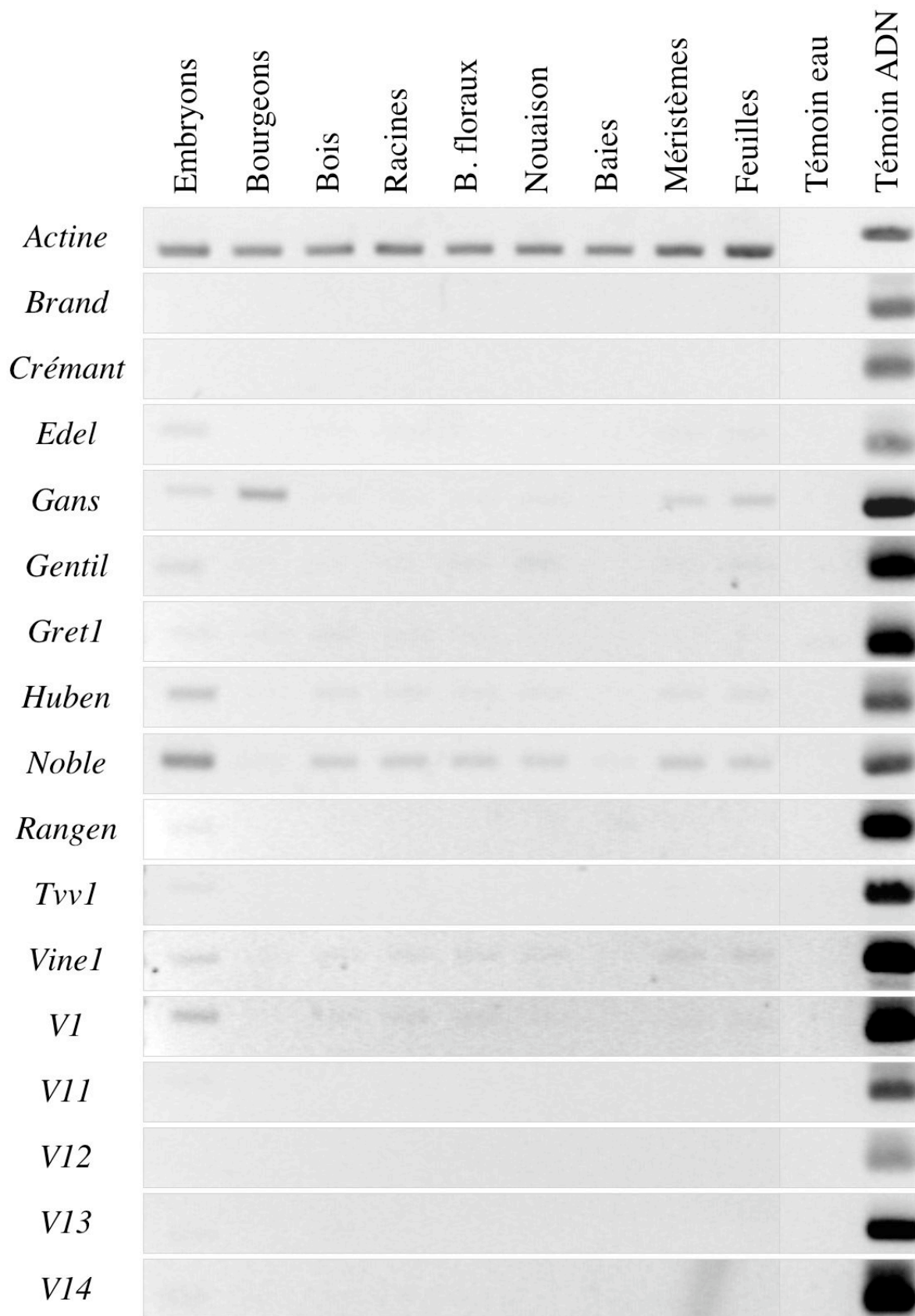


Figure 44 : Détection par RT-PCR de transcrits spécifiques de 16 rétrotransposons dans différents organes en conditions normales de croissance.

permis l'amplification des familles *Edel*, *Gans*, *Gentil*, *Huben*, *Noble*, *Vine-1* et *VI*. Des signaux très faibles ont été détectés dans ces embryons pour les éléments *Gret1* et *Tvv1*. Lorsque l'on considère les extractions réalisées à partir de bourgeons, l'élément *Gans* donne un signal alors qu'aucune autre famille de rétrotransposons n'est amplifiée. Aucune amplification n'a été obtenue à partir des ADNc de pellicules de baies. Des transcrits appartenant aux familles *Gans*, *Noble* et *Vine-1* ont été détectés dans les ADNc de méristèmes apicaux et de feuilles. Finalement, la présence de transcrits de la famille *Noble* a aussi été détectée dans le bois, les racines, les boutons floraux et les baies à la nouaison, et de faibles amplifications ont été obtenues à partir de ces 4 organes pour les familles *Huben*, *Vine-1* et *VI*.

Légende de la Figure 44 : Embryons : embryons somatiques. Bourgeons : bourgeons avant débourrement. B. floraux : boutons floraux. Nouaison : baies à la nouaison. Baies : pellicules de baies mûres. Méristèmes : méristèmes apicaux. Témoin eau : témoin d'amplification sans ADN. Témoin ADN : témoin d'amplification sur ADNg de Riesling.

II.3.2 - Evaluation de l'influence de deux stress sur le niveau de transcription dans les feuilles

a) Blessure mécanique

Gènes témoins de stress

Les résultats présentés dans la Figure 45 montrent que l'amplification des transcrits par le couple d'amorces *PR1* a été obtenue dans toutes les conditions, mais qu'un signal de plus forte intensité a été obtenu à partir des feuilles non-blessées et à partir des feuilles prélevées 24h et 48h après la blessure de la série n°1. Si la tendance a été sensiblement la même dans les deux répétitions, l'augmentation de l'intensité du signal entre les feuilles prélevées à 6h et 12 h d'une part, et les autres temps d'autre part, était plus visible dans la première série. Pour le couple d'amorces *Def1*, des quantités plus importantes de transcrits ont été détectées dans les feuilles prélevées 12h après la blessure de la série 1 et dans les feuilles saines de la série 2. Cependant, l'intensité des signaux n'était pas comparable dans les deux séries puisque les amplifications *Def1* obtenues dans la deuxième série étaient globalement plus fortes. Dans les

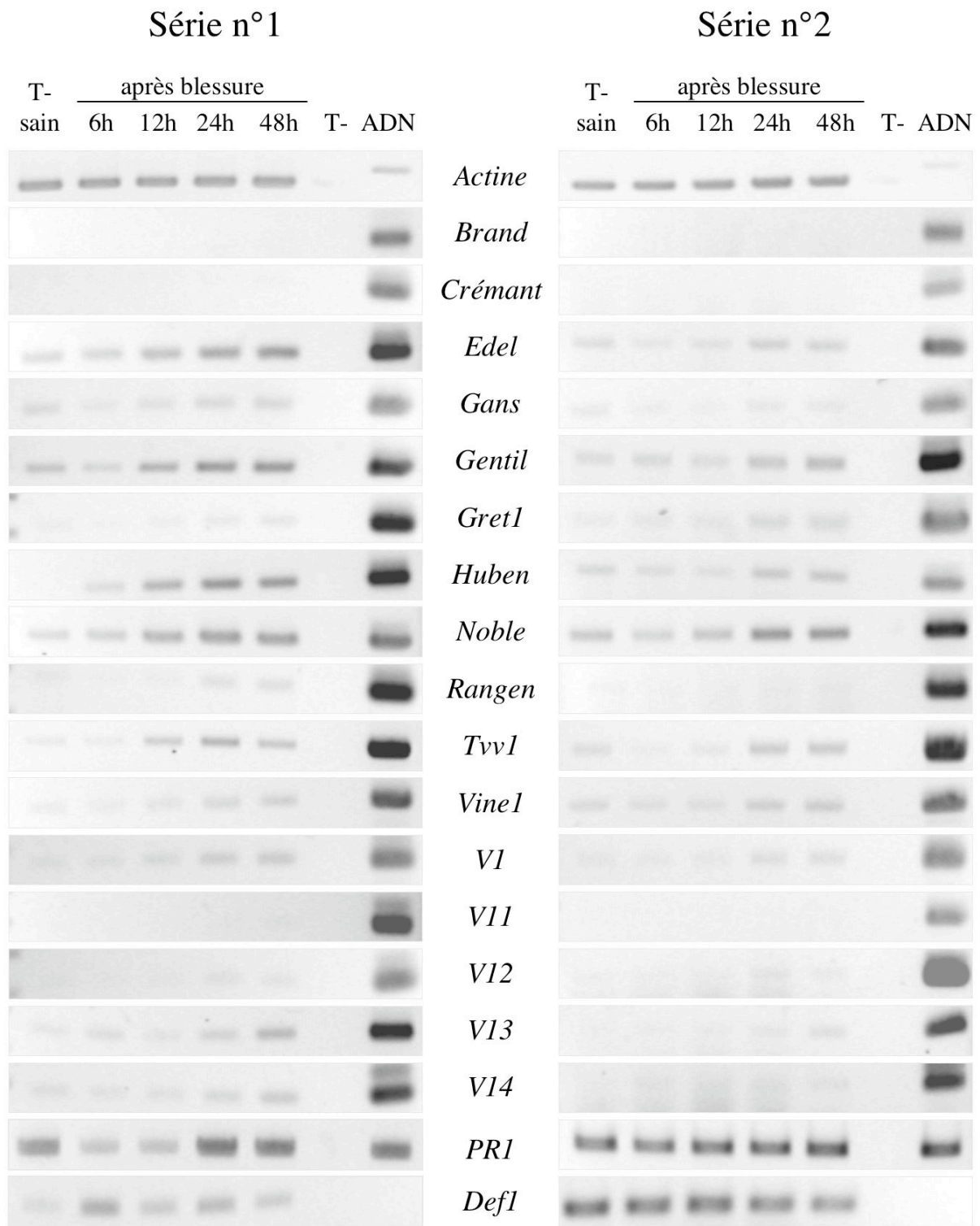


Figure 45 : Evaluation par RT-PCR de l'influence de l'induction par la blessure mécanique sur le niveau de transcription de 16 rétrotransposons dans les feuilles.

deux répétitions, les amplifications *Def1* étaient plus faibles dans les feuilles prélevées 24h après la blessure.

Transcription des rétrotransposons

Les résultats présentés dans la Figure 45 montrent que les ADNc synthétisés à partir d'extraits d'ARN de feuilles ne permettent pas l'amplification de transcrits pour les familles *Brand*, *Crémant* et *VII*, que ce soit avant ou après l'application de la blessure mécanique.

Lorsque l'on considère les ADNc synthétisés à partir des ARN extraits de feuilles non-blessées, les éléments *Huben*, *Tvv1*, *Vine-1* et *V13* ne sont pas amplifiés, et des signaux faibles sont obtenus pour les familles *Edel*, *Gans*, *Gentil*, *Noble*, *VI* et *VI4*. Pourtant, les niveaux d'amplification de ces 10 familles sont plus intenses lorsque les amplifications sont réalisées à partir d'échantillons d'ADNc de feuilles blessées. Les signaux augmentent nettement 12h après la blessure, atteignent leur maximum à 24h et restent stables après 48h. Cette tendance a été confirmée dans les deux répétitions, malgré le fait que les signaux étaient moins intenses dans la seconde série. De la même manière, alors qu'aucun signal n'a été détecté pour les feuilles non-blessées, une amplification faible des éléments *Gret1*, *Rangen* et *VI2* a été observée dans les ADNc de feuilles prélevées 24h et 48h après la blessure.

Légende de la Figure 45 : Séries n°1 et 2 : répétitions de l'expérience sur des feuilles et des plantes indépendantes. T- sain : témoin sans blessure. Après blessure : feuilles prélevées à différents temps après la blessure. T- : témoin d'amplification négatif sans ADN ni ADNc. ADN : témoin d'amplification sur ADN génomique de Riesling.

b) Inoculation par le mildiou

Témoin d'infection

Afin de s'assurer que l'inoculation par le mildiou a été effective sur les feuilles de Riesling à partir desquelles les ARN ont été extraits, nous avons vérifié le résultat de l'infection des feuilles de Muscat ottonel, la variété sensible, par la même solution de sporanges de mildiou. La Figure 46 montre que, 6 jours après la pulvérisation de la solution de sporanges, un voile blanc de mycélium était visible sur toutes les feuilles témoins de Muscat ottonel. Cependant, ce voile ne couvrait pas toute la surface des feuilles et des zones non-infectées étaient visibles.



Figure 46 : Observation d'une feuille témoin de Muscat ottonel, 6 jours après l'inoculation de *P. viticola*.

Gènes témoins

Malgré le soin pris pour l'étalonnage des quantités d'ADNc, le niveau d'amplification des transcrits du gène de l'actine s'est avéré plus fort pour certains échantillons : à partir des ADNc de feuilles saines dans les séries n°1 et 2, et à partir des ADNc de feuilles prélevées 12h après l'inoculation, dans la série n°2 (Figure 47). Le niveau de transcription des familles de gènes témoins *PRI* et *def1* n'a pu être estimé dans la première série de feuilles inoculées et les échantillons n'étaient malheureusement plus disponibles en quantité suffisante pour répéter les amplifications PCR. Les signaux obtenus par l'amplification des ADNc correspondant à la famille de gènes *PRI* semblent évoluer dans le même sens que les résultats obtenus pour l'actine, mais le signal *PRI* est plus fort lorsque l'amplification est réalisée à partir d'ADNc de feuilles prélevées 48h après l'infection.

Dans le cas du couple d'amorces *Def1*, un signal plus intense a été observé pour les amplifications réalisées à partir des ADNc de feuilles saines et de feuilles prélevées 6h après l'inoculation. L'amplification des ADNc correspondant aux familles de gène *Def1* est moins forte dans les feuilles prélevées après 24h et 48h.

Transcription des rétrotransposons

Les résultats présentés dans la Figure 47 montrent que les 16 rétrotransposons ont été peu amplifiés dans les feuilles après l'inoculation par le mildiou. Les transcrits des éléments *Brand*, *Crémant*, *Gret1*, *Rangen*, *V1*, *V11* et *V12* n'ont pas été détectés dans ces conditions. Les transcrits des familles *Edel*, *Gans*, *Gentil*, *Huben*, *Noble*, *Tvv1*, *Vine-1* et *V13* ont été peu ou pas détectés dans les feuilles saines, alors qu'un signal d'amplification plus important a été obtenu à partir des ADNc de feuilles prélevées 12h après l'application du stress. Les transcrits des familles *Edel* et *Noble* ont été les plus amplifiés dans les feuilles inoculées. De manière générale, les amplifications obtenues dans la deuxième série d'inoculation par le mildiou sont plus intenses et semblent suivre la tendance du signal d'amplification de l'actine. Les transcrits des rétrotransposons sont, dans l'ensemble, amplifiés plus fortement à partir des ADNc de feuilles prélevées 48h après l'infection, alors que le signal de l'actine est plus faible, ce qui semble indiquer que la quantité relative des transcrits de rétrotransposons est plus importante dans ces feuilles. L'élément *V14* présente quant à lui un profil différent dans chacune des deux séries puisqu'il est détecté dans les ADNc des feuilles prélevées 6h après induction dans la première série, alors qu'il n'est détecté que dans les feuilles saines de la seconde répétition.

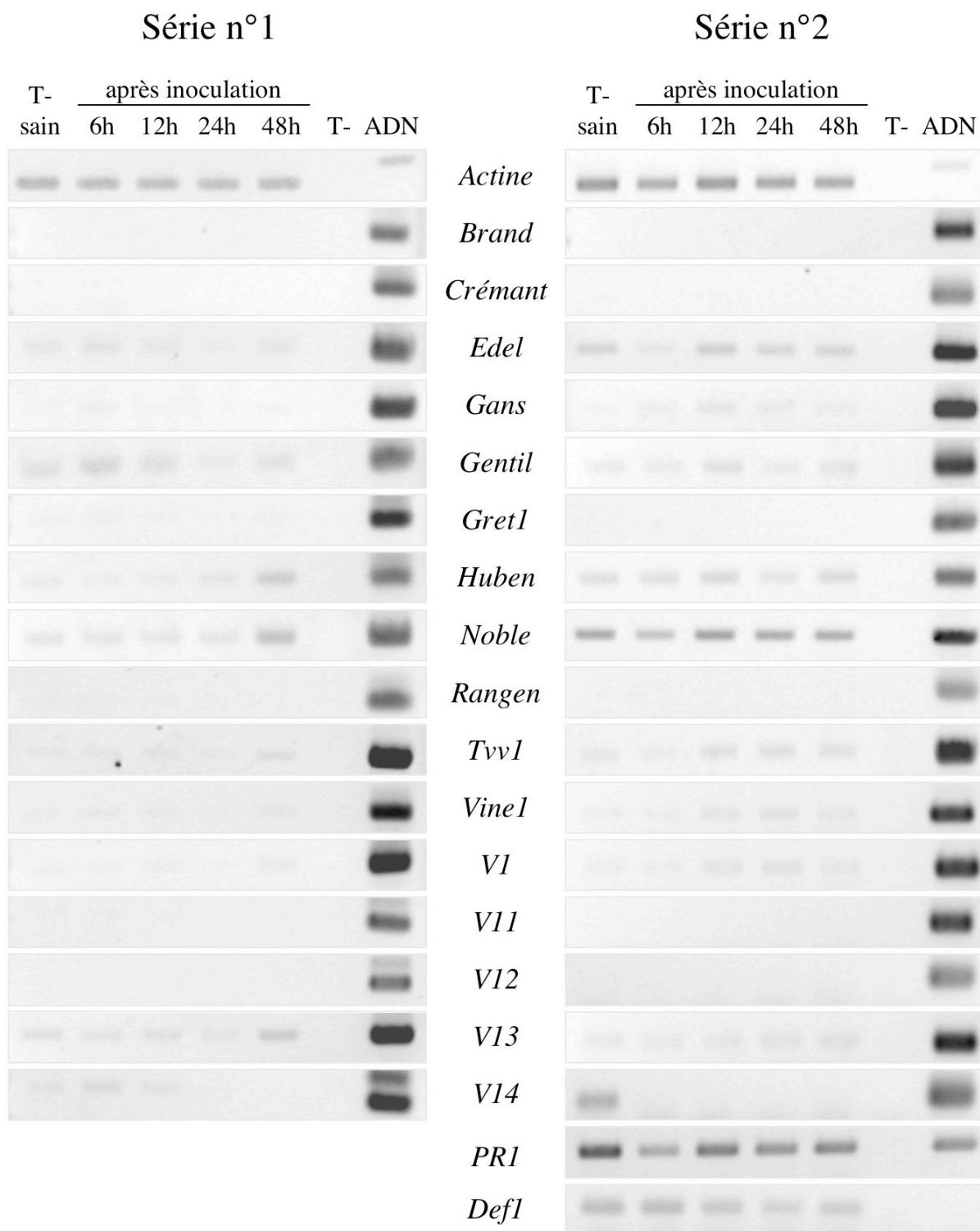


Figure 47 : Evaluation de l'influence de l'induction par l'inoculation de mildiou sur le niveau de transcription de 16 rétrotransposons dans les feuilles.

Légende de la Figure 47 : Séries n°1 et 2 : répétitions de l'expérience sur des feuilles et des plantes indépendantes. T- sain : témoin sans mildiou. Après inoculation : feuilles prélevées à différentes temps après la pulvérisation d'une solution de sporange de mildiou sur la face inférieure des feuilles. T- : témoin d'amplification négatif sans ADN ni ADNc. ADN : témoin d'amplification sur ADN génomique de Riesling.

II.3.3 - Etude de la variabilité des séquences transcrites

a) Variabilité de la région gag des transcrits

Les produits RT-PCR, amplifiés à partir des extractions d'ARN de feuilles de Riesling prélevées 24h après la blessure par chaque couple d'amorces ciblant la région *gag* des rétrotransposons, ont été clonés. Pour chaque rétrotransposon, quatre ADNc clonés ont été séquencés, puis leurs séquences ont été comparées pour évaluer la variabilité des transcrits de chaque famille.

Aucune séquence ADNc n'a été amplifiée à partir des ADNc de feuilles pour les familles *Brand*, *Crémant* et *Rangen*. Les résultats présentés dans la Figure 48 montrent que des fragments très proches des tailles attendues ont été amplifiés spécifiquement par les couples d'amorces dessinés pour les autres familles. Les séquences obtenues pour chaque famille diffèrent les unes des autres par de rares substitutions de bases ou des petites délétions d'une ou deux bases, sauf dans le cas de la famille *VI3* dont toutes les séquences ADNc sont strictement identiques. Pour chacune des familles *Gret1*, *Noble*, *Tvv1*, *Vine-1*, *VI* et *VI3*, les quatre séquences sont de tailles identiques. Les séquences appartenant à la même famille de rétrotransposons présentent des taux d'identité de 83,1 à 100%. Les séquences les plus divergentes ont été obtenues pour les familles *Huben*, *Tvv1*, *VII* et *VI2*.

Les différences observées entre les séquences obtenues pour chaque rétrotransposon peuvent avoir plusieurs origines. Tout d'abord, les ADNc ont pu être synthétisés à partir de transcrits différents, ce qui signifie que plusieurs copies de la famille du rétrotransposon sont exprimées ou que les transcrits de la même copie diffèrent par l'apparition de mutations lors de l'étape de transcription. A l'inverse, les séquences ADNc sont peut-être issues des transcrits de la même copie exprimée mais des mutations sont apparues pendant les phases d'amplification

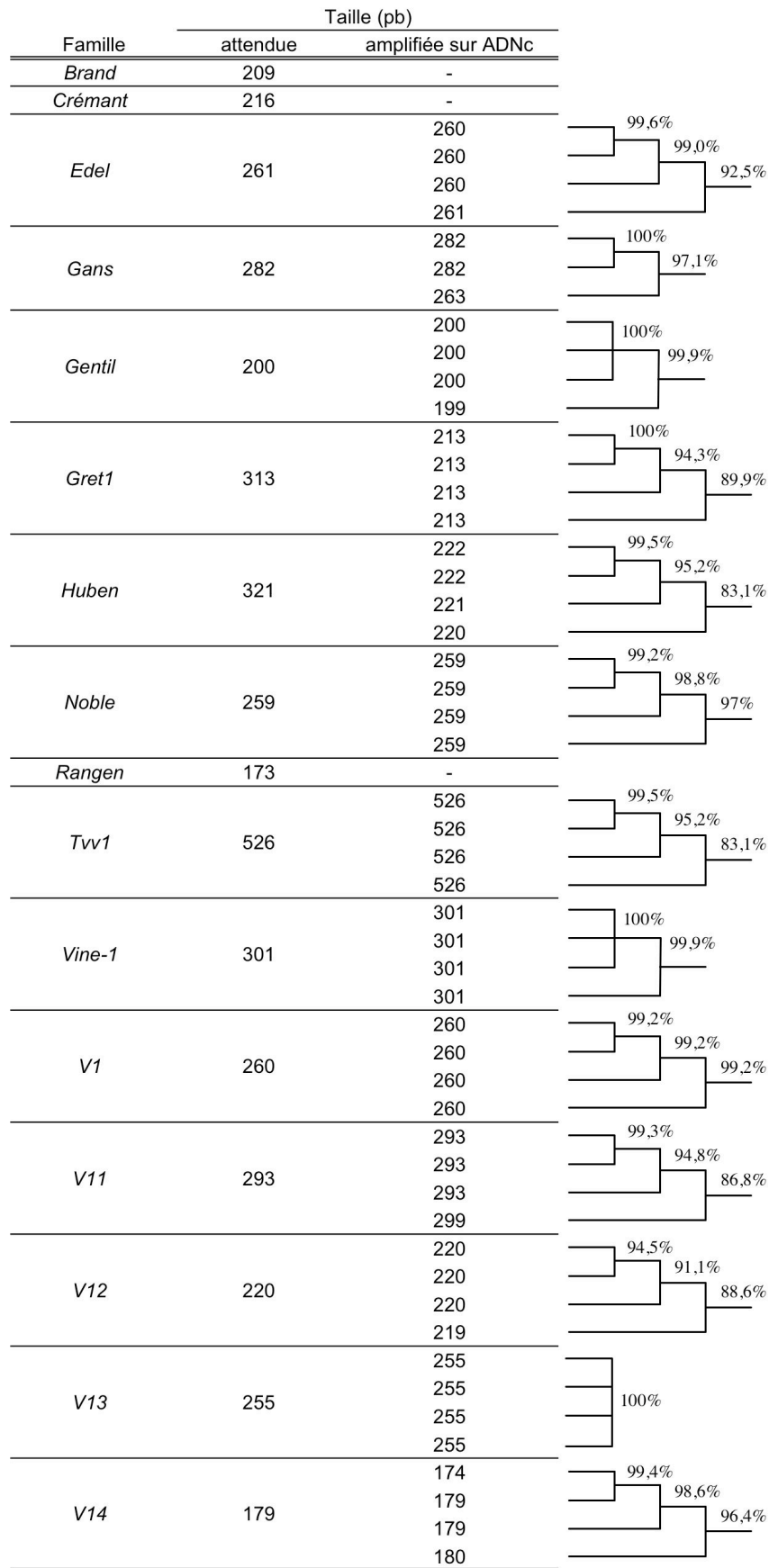


Figure 48 : Comparaison des séquences ADNc amplifiées par chaque couple d'amorces de rétrotransposon.

RT-PCR, de clonage ou de séquençage. Nous avons en effet estimé que notre protocole de PCR-clonage-PCR-séquençage introduit environ 1 mutation sur 770 pb séquencées (Moisy *et al.*, 2008).

b) Variabilité de la région UTL des transcrits *Tvv1*

Le séquençage des produits RT-PCR amplifiés avec le couple d'amorces encadrant la région UTL des éléments de la famille *Tvv1* nous a permis d'obtenir 10 séquences correspondant à la région UTL de copies *Tvv1* exprimées 24h après la blessure dans les feuilles de Riesling. Ces séquences ont été comparées aux séquences UTL de copies *Tvv1* que nous avons identifiées dans les génomes du Riesling, de PN40024, et aux séquences UTL de deux copies *Tvv1* particulières : la copie *Tvv1*-VB du génome de la variété Vidal blanc, que nous avons décrite comme la première copie complète de la famille *Tvv1* ; et l'UTL d'une copie délétée du cabernet sauvignon, *Tvv1*-Δ3001 (ces deux éléments sont décrits dans l'article Moisy *et al.*, 2007, présenté dans la Partie IV).

Les résultats présentés dans la Figure 49 montrent que les séquences UTL obtenues à partir d'ADNc de feuilles de Riesling blessées correspondent à 6 copies différentes, puisque les séquences Riexp10, Riexp08, Riexp01, Riexp07, Riexp02 et 05, et Riexp04 correspondent respectivement à des copies *Tvv1* dont les régions UTL font 740, 784, 823, 843, 850 et 890 pb. Les UTL de ces copies se différencient par des mutations ponctuelles, mais leurs tailles varient principalement par des insertions/délétions de plusieurs dizaines de bases. Les UTL Riexp10 et Riexp07 sont respectivement très similaires en taille et en séquence aux UTL des copies *Tvv1*-PN140 et *Tvv1*-PN141 identifiées dans le génome de PN40024.

Les UTL obtenues à partir d'ADNc se positionnent dans l'arbre à proximité de séquences UTL appartenant à des copies dont les LTR 5' et 3' ont peu divergé, donc des copies qui se sont insérées relativement récemment dans le génome de la vigne. Ce résultat semble confirmer l'hypothèse selon laquelle des copies *Tvv1* actives et autonomes sont présentes dans le génome de la vigne. La répartition des séquences UTL en fonction de l'âge des copies montre que les copies transcrites sont des copies « jeunes » qui ont des séquences UTL très proches, alors que les copies plus anciennes ne sont pas exprimées et leurs UTL sont très différentes les unes par rapport aux autres. Ce résultat semble indiquer que les régions UTL évoluent et divergent les unes des autres une fois que la copie est intégrée au génome, et non pendant le cycle de rétrotransposition.

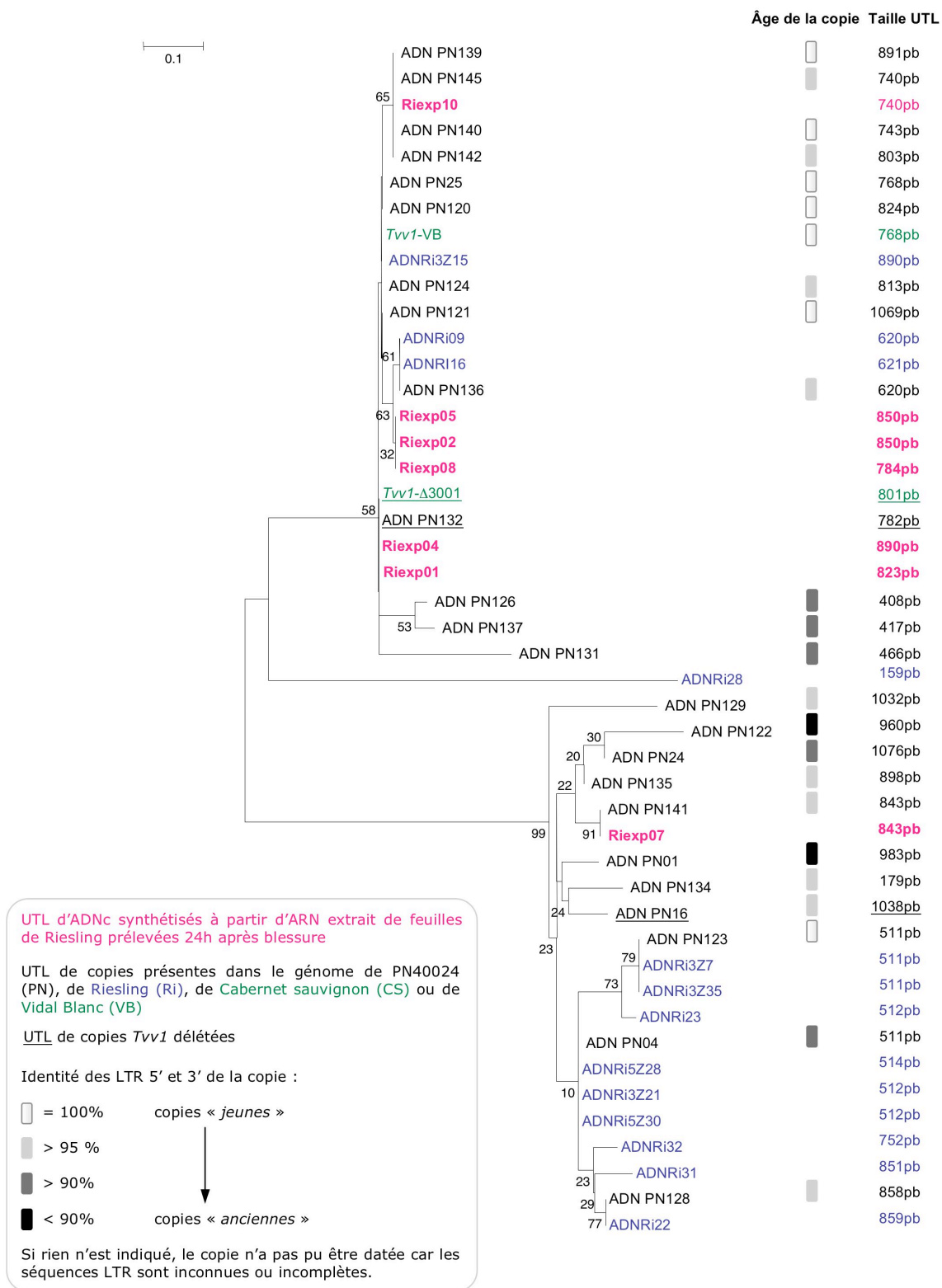


Figure 49 : Arbre Neighbor-Joining construit à partir des séquences UTL de copies *Tvv1* transcrites dans les feuilles Riesling après la blessure, et de copies identifiées dans les génomes de Riesling, de PN40024, de Vidal blanc et de Cabernet sauvignon.

II.4 - Discussion des résultats de l'analyse du niveau de transcription par RT-PCR

a) *Transcription des rétrotransposons dans les organes*

Régulation de l'expression en conditions normales de croissance

L'évaluation par RT-PCR semi-quantitative du niveau de transcription des rétrotransposons dans les différents organes végétatifs et reproducteurs de la vigne montre que chaque famille présente un profil d'expression particulier. Tout d'abord, des transcrits n'ont été détectés que pour 9 familles sur les 16 testées. Les éléments *Brand*, *Crémant*, *Rangen*, *V11*, *V12*, *V13* et *V14* n'ont été détectés dans aucun organe, alors que les éléments *Edel*, *Gans*, *Gentil*, *Gret1*, *Huben*, *Noble*, *Tvv1*, *Vine-1* et *V1* sont transcrits dans au moins un type de tissu. Ces résultats semblent indiquer que l'expression des rétrotransposons est régulée dans les organes en conditions normales de croissance.

Confirmation des résultats de la recherche d'EST

Ces conclusions sont en accord avec les résultats obtenus lors de l'analyse des séquences EST de vigne puisque les rétrotransposons qui n'ont pas été détectés par RT-PCR ne présentaient pas ou peu de séquences EST dans les banques disponibles pour la vigne.

L'élément *V14* est le seul rétrotransposon qui donne des résultats différents dans ces deux études, puisqu'il n'a été amplifié par RT-PCR dans aucun organe alors que de nombreuses séquences EST lui correspondent dans les bases de données. Cependant, 80% de ces séquences EST ont été obtenues à partir de feuilles de Chardonnay soumises à une condition de stress abiotique, ce qui peut vouloir dire que l'élément *V14* n'est pas exprimé dans les organes en conditions normales de croissance (sans stress), ou que le génome de Chardonnay contient des copies *V14* actives, mais pas celui du Riesling.

Profil d'expression particulier de la famille Noble

L'élément *Noble* présente un profil particulier puisqu'il a été détecté par RT-PCR dans presque tous les organes dans lesquels ses transcrits ont été recherchés. Il est également l'élément dont les séquences EST sont représentées dans le plus grand nombre de conditions. Ces deux résultats complémentaires semblent indiquer que l'élément *Noble* est transcrit de manière constitutive dans presque tous les organes de la vigne en conditions normales de

croissance. Cette famille présente donc un profil d'expression particulier puisque chez les plantes, l'expression des rétrotransposons semble au contraire très régulée et spécifique de certains tissus, comme le prouve l'élément *Tnt1* du tabac qui ne s'exprime que dans les racines (Pouteau *et al.*, 1991), les éléments *Tto1* du tabac et *Tos17* du riz qui sont transcrits uniquement dans les cellules en culture, les éléments *Tto1*, *Tos10* et *Tos17* qui ne sont pas exprimés dans les feuilles (Hirochika, 1993; Hirochika *et al.*, 1996), l'élément *CIRE1* de l'oranger qui est exprimé dans les racines, mais pas dans les feuilles (Rico-Cabanas and Martinez-Izquierdo, 2007) ou encore l'élément *PREM-2* du maïs qui s'exprime lors les premiers stades du développement des microspores, mais ne s'exprime ni dans le pollen, ni dans les autres tissus (Turcich *et al.*, 1996).

Le profil d'expression de la famille *Noble* chez la vigne se rapproche plutôt de celui qui a été décrit pour l'élément *Ogre* qui s'exprime dans les racines, les feuilles et les fleurs chez le pois (Neumann *et al.*, 2003).

Pas de transcription dans les pellicules de baies

Aucune famille n'est transcrite dans les pellicules de baies de Riesling et les EST de rétrotransposons trouvées dans des banques obtenues à partir de baies sont très peu nombreuses. Ces résultats suggèrent que les rétrotransposons s'expriment rarement dans ce type de tissu. Compte tenu du fait qu'aucune famille ne s'exprime dans les pellicules de baies, il est possible qu'un mécanisme de régulation bloque de manière non-spécifique l'expression des rétrotransposons, et probablement celle d'autres gènes. C'est par exemple le cas lorsque la méthylation de l'ADN est plus forte dans certaines cellules et qu'elle provoque un phénomène de régulation généralisé à la région chromosomique méthylée (Okamoto and Hirochika, 2001; Madlung and Comai, 2004).

Transcription importante dans les embryons somatiques

Des transcrits correspondant à 9 familles de rétrotransposons ont été détectés dans les cals générés à partir de l'embryogenèse somatique de culture de filet d'anthères. Les rétrotransposons de la vigne s'expriment donc plus fortement dans ce type de tissu. Il a été démontré que les changements provoqués par la culture de cellules sont parfois associés à des mécanismes mutagènes et à l'activation de rétrotransposons. Chez l'orge, l'élément *BARE-1* est exprimé dans les cals embryogènes et dans les feuilles régénérées à partir de protoplastes (Suoniemi *et al.*, 1996). Plusieurs études prouvent en effet que la transcription des rétrotransposons induite par la culture de tissus peut être suivie de l'insertion de nouvelles

copies dans le génome. De nouvelles insertions de *Tnt1* ont été détectées dans 25% des plantes régénérées à partir de protoplastes (Melayah *et al.*, 2001). De la même manière, le nombre de copies de *Tto2* et *Tnt1* est plus important dans les cellules re-différenciées, et dans le cas de l'élément *Tto1*, le nombre de copies est jusqu'à 10 fois supérieur dans les cellules et les plants de tabac régénérés. Chez le riz, les niveaux de transcription des trois éléments *Tos10*, *Tos17* et *Tos19* sont également plus importants lors de la culture de tissus, et entre 5 et 30 nouvelles copies de *Tos17* ont été détectées dans les plantes régénérées par ce procédé, notamment dans le gène *phytochrome A* et dans un gène de la famille des récepteurs kinase (Hirochika *et al.*, 1996).

Le taux de transcription des rétrotransposons détecté dans les cals embryogènes de vigne peut donc avoir des conséquences importantes sur le génome de la plante. Si le taux de rétrotransposition est proportionnel au niveau transcription, le génome de la future plante pourrait être affecté par de nouvelles mutations d'insertions et la probabilité d'observer de la variabilité somaclonale peut s'en trouver considérablement accrue.

b) Influence des stress sur le niveau de transcription dans les feuilles

B1) Stress abiotique : la blessure

Les niveaux d'amplification des rétrotransposons à partir d'ADNc des feuilles témoins, non-blessées et non-infectées étaient concordants avec les résultats précédemment obtenus sur les ADNc de feuilles en conditions normales de croissance. Dans le cas de l'induction par la blessure mécanique, il est probable que la transcription des gènes de la famille *PR1* augmente entre 12h et 48h. L'augmentation de la transcription des gènes de la famille *Def1* semble plus rapide en réponse à l'induction et diminue après 24h. Ces deux observations confirment que la blessure a bien eu un effet inducteur sur l'expression des gènes de réponse au stress.

Familles non transcrites

Malgré l'application de la blessure, des transcrits correspondant aux familles *Brand*, *Crémant*, et *VII* n'ont pas été détectés. Ce résultat, corrélé à leur très faible présence dans les banques de séquences EST, confirme que ces trois familles ne sont vraisemblablement pas transcrites dans les feuilles de vigne et que la blessure n'induit pas leur expression. Comme nous l'avons

vu précédemment, aucune des copies des familles *Brand* et *VII* détectées dans le génome de PN40024 ne semble correspondre à un élément autonome. L'évaluation de leur niveau de transcription chez le Riesling confirme que ces deux familles ne sont probablement pas actives. A l'inverse, le génome de PN40024 contient des copies de la famille *Crémant* qui présente les caractéristiques de copies autonomes, mais aucun transcrit n'a été détecté chez le Riesling. La présence de 5 séquences EST dans une banque construite à partir de feuilles de Chardonnay soumises à un stress abiotique (non précisé) semble indiquer que cette famille est transcrite dans cette condition très particulière, mais pas en réponse à la blessure.

Induction de la transcription

Pour les autres familles, l'application de la blessure est presque systématiquement suivie par une augmentation du niveau de transcription. Dans le cas des familles *Gans*, *Gret1*, *Rangen*, *Vine-1*, *VI*, *VI3* et *VI4*, les niveaux de transcription sont très faibles dans les feuilles non-blessées, mais semblent augmenter 24h ou 48h après la blessure. L'expression des familles *Edel*, *Gentil*, *Huben*, *Noble* et *Tvv1* est, quant à elle, nettement plus forte après l'induction. Il est possible que l'augmentation de la présence de transcrits soit le résultat de l'augmentation de l'activité transcriptionnelle globale de la cellule, et non pas d'une induction spécifique. Mais dans un tel cas de figure, tous les gènes de l'hôte seraient également surreprésentés, la proportion de rétrotransposons resterait inchangée et l'augmentation des transcrits de rétrotransposons ne serait probablement pas visible. Le fait que la quantité de transcrits des rétrotransposons soit généralement plus importante après la blessure peut vouloir dire que l'induction ne cible pas spécifiquement une famille en particulier, mais agit sur presque tous les rétrotransposons de manière uniforme, comme dans le cas d'une réponse épigénétique.

Compte tenu du fait que le génome de PN40024 contient beaucoup de copies des familles *Edel* et *Noble*, que certaines d'entre elles sont potentiellement autonomes, et que ces familles sont transcrites chez les Riesling, particulièrement en réponse à la blessure, ces deux familles sont très certainement actives chez la vigne.

En ce qui concerne les familles *Gans*, *Gentil*, *Gret1*, *Huben*, *Tvv1*, *Rangen*, *Vine-1*, *VI*, *VI3* et *VI4*, il n'est pas possible d'affirmer que des copies sont actives et autonomes dans le génome de la vigne car, bien qu'elles soient transcrites dans les feuilles de Riesling en réponse à la blessure, aucune de ces familles ne présente de copie potentiellement autonome dans le génome de PN40024. D'autres expériences seront nécessaires pour savoir si de telles copies ont été éliminées du génome de PN40024 lors des cycles d'autofécondation ou si ces familles sont effectivement non-autonomes.

B2) Stress biotique : l'infection par le mildiou

Infection hétérogène sur la variété témoin

Les feuilles de Riesling sur lesquelles a été pulvérisée la solution de sporanges de mildiou ayant été prélevées au maximum 48h après l'infection, le délai était trop court pour permettre l'apparition des premiers symptômes visibles. Des feuilles de Muscat ottonel, une variété très sensible à ce pathogène, ont été infectées par la même solution pour vérifier l'agressivité du pathogène et l'efficacité de l'inoculation. Ces feuilles, observées 7 jours après la pulvérisation de l'inoculum, présentaient toutes un voile discontinu de mycélium sur leur face inférieure, ce qui confirme l'efficacité de l'infection mais indique que le stress a été appliqué de manière hétérogène sur la surface des feuilles.

Il est donc probable que tous les échantillons de feuilles de Riesling infectées n'aient pas été soumis à la même pression du pathogène et donc au même niveau de stress. A l'inverse, la blessure mécanique, appliquée à l'aide d'une plaque munie de pointes, était un stress régulier, très homogène et facilement répétable.

Cette hétérogénéité dans l'application du stress biotique a certainement été en partie pondérée par le fait que les échantillons de feuilles prélevées à un temps donné ont été mélangés avant d'en extraire les ARN. Cependant, les effets, sur les résultats de l'étude, d'une trop grande variation du stress au départ, n'ont certainement pas été totalement contrebalancés par cette seule précaution.

L'utilisation de sachets plastique pour maintenir une humidité suffisante au développement du mildiou n'a vraisemblablement pas eu d'effet significatif sur l'évaluation du niveau de transcription puisque les résultats obtenues à partir des ADNc de feuilles saines (mais conservées 12h dans le sachet), de feuilles prélevées 12h et 24h après l'inoculation, sont équivalents.

Induction de la transcription des rétrotransposons par l'infection

Les éléments *Edel*, *Gentil*, *Huben*, *Noble*, *Tvv1* et *V13* apparaissent plus fortement transcrits après l'induction par l'infection au mildiou, comme c'est le cas pour le rétrotransposon *Tnt1* du tabac en réponse à l'attaque de pathogènes ou à l'utilisation d'éliciteurs fongiques (Grandbastien *et al.*, 1997; Melayah *et al.*, 2001), ou chez le riz pour l'élément *Rim2* en réponse à l'infection par *Magnaporthe grisea* (He *et al.*, 2000). Dans le cas de l'élément *V13*,

dont les transcrits n'ont été détectés ni dans les banques EST, ni dans les organes, sa transcription en réponse à l'infection par le mildiou peut signifier que son expression est restreinte à certaines conditions bien particulières.

L'expression des gènes de la plante en réponse à l'attaque d'un pathogène est généralement induite par une cascade de transduction de signaux qui déclenche finalement la réponse de l'organisme. Il a été démontré que des motifs particuliers, présents dans les LTRs de rétrotransposons, sont susceptibles de participer à l'induction de l'expression de ces éléments en réponse à un stress, notamment suite à l'attaque d'un pathogène (Suoniemi *et al.*, 1996; Casacuberta *et al.*, 1997; Grandbastien *et al.*, 1997; Grandbastien, 1998; Takeda *et al.*, 1999). L'augmentation du niveau de transcription des rétrotransposons en réponse à la blessure et à l'attaque par le mildiou ne permet pas d'exclure l'hypothèse selon laquelle des mécanismes similaires sont impliqués dans la régulation de l'expression des rétrotransposons chez la vigne. La recherche de motifs comparables à ceux qui ont été décrits chez *BARE1*, *Tnt1* et *Tto1*, dans les LTRs des familles dont la transcription est induite par l'infection au mildiou, devrait nous permettre de savoir si ces familles sont susceptibles d'être induites en réponse à l'infection du mildiou par ce même type de mécanisme.

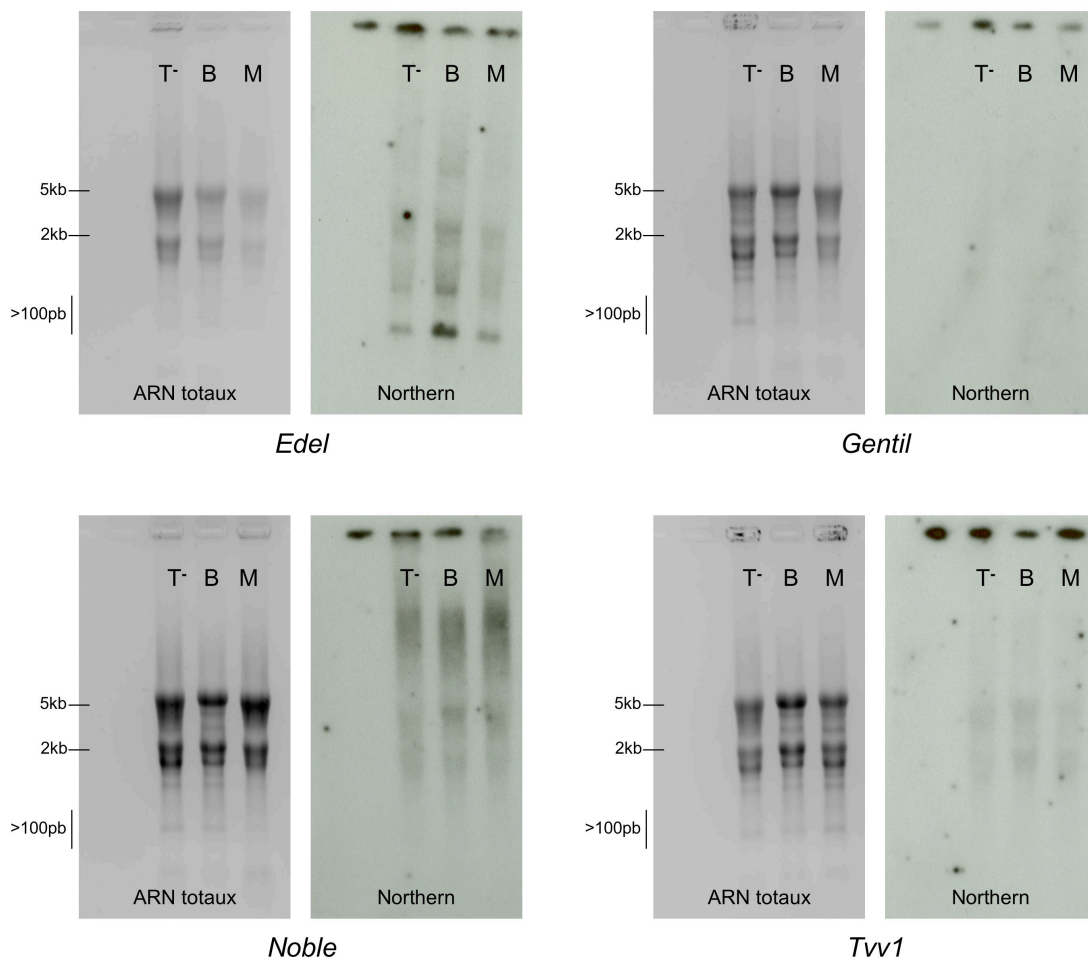
Chapitre III : Evaluation de la taille des transcrits

III.1 - Détection des transcrits de 4 rétrotransposons par Northern

Les couples d'amorces utilisés pour la RT-PCR ont été choisis dans ou à proximité du domaine *gag* pour permettre d'amplifier spécifiquement les ADNc correspondant aux transcrits d'une famille de rétrotransposons. La séquence des amorces devait permettre d'amplifier le maximum de copies de chaque famille, mais des transcrits de copies ayant subis un grand nombre de mutations n'ont peut-être pas pu être amplifiés. De plus, il ne faut pas exclure l'hypothèse selon laquelle certains transcrits partiellement dégradés ou incomplets n'ont pas été détectés par cette méthode. En effet, les ADNc ont été synthétisés à partir de la queue poly(A) des ARN messagers, alors que le domaine *gag* est plus proche de l'extrémité 5'. Il est donc possible que des ADNc incomplets aient été synthétisés lors de l'étape de transcription inverse, notamment dans le cas des ARNm de grands rétrotransposons. Ainsi, il est probable que le niveau de transcription de certaines familles ait été sous-évalué, comme par exemple pour *Gret1* qui n'a été détecté que très rarement par RT-PCR. Pour répondre à cette question et estimer la taille des transcrits de certains rétrotransposons, une expérience de Northern blot a été réalisée.

Nous avons recherché les transcrits correspondant aux familles *Edel*, *Gentil*, *Tvv1* et *Noble* par Northern blot parmi des ARN totaux extraits à partir de feuilles de Riesling saines, prélevées 24h après une blessure ou prélevées 48h après l'inoculation de mildiou. Ces quatre rétrotransposons ont été choisis car l'évaluation de leur niveau de transcription par RT-PCR a révélé qu'ils présentaient les niveaux d'expression les plus forts en réponse au stress, ce qui faisait d'eux de bons candidats pour la détection des transcrits par Northern.

Les stress et les extractions d'ARN ont été réalisés dans les mêmes conditions que précédemment. Pour chaque condition, 10 µg d'ARN totaux ont été séparés par électrophorèse sur gel d'agarose en conditions dénaturantes puis transférés sur des membranes de nylon (Voir Partie V : Matériels et Méthodes). Des sondes ADN ont été



ARN totaux : observation des ARN totaux transférés par coloration de la membrane au BET. Northern : Films autoradiographiques exposés après 7 jours d'exposition aux membranes hybridées avec les sondes marquées. T- : ARN extraits de feuilles saines. B : ARN extraits de feuilles prélevées 24h après la blessure. M : ARN extraits de feuilles prélevées 48h après inoculation par le mildiou.

Figure 50 : Evaluation de la taille des transcrits des rétrotransposons *Edel*, *Gentil*, *Noble* et *Tvv1* par Northern blot.

synthétisées par PCR, à partir d'ADN génomique de Riesling, en utilisant les couples d'amorces dessinés pour la RT-PCR qui ciblent la région *gag* de chaque élément, et elles ont ensuite été marquées radioactivement (Voir Partie V : Matériels et Méthodes). Chacune des quatre sondes a été hybridée à une membrane indépendante, portant les ARN obtenus dans les trois conditions (feuilles saines, blessée et infectée). Un film autoradiographique a finalement été exposé pendant 7 jours aux radiations émises par les sondes hybridées sur les membranes.

III.2 - Résultats

Les résultats présentés dans la Figure 50 montrent que tous les ARN ont été correctement transférés sur la membrane et qu'ils n'étaient pas dégradés puisque les bandes d'ARN ribosomiaux étaient bien visibles. Le marqueur de taille ARN utilisé dans cette expérience n'ayant pas été révélé correctement, la taille des transcrits a été estimée en comparant les bandes d'ARN ribosomiaux aux marqueurs présents sur d'autres gels dénaturant réalisés avec des ARN de feuille de vigne.

Aucun signal n'a été détecté suite à l'hybridation de la sonde *Gentil* sur les ARN totaux.

Des signaux, correspondant probablement à l'hybridation non-spécifique des sondes aux ARN ribosomiaux, ont été détectés pour les éléments *Noble* et *Tvv1*.

Dans le cas de l'élément *Edel*, des transcrits de quelques centaines de paires de bases ont été détectés dans les trois conditions. L'intensité du signal était cependant plus forte dans les ARN de feuilles blessées bien que la quantité d'ARN transférée semblait plus faible que pour les autres conditions. De la même façon, un signal de la même intensité a été observé entre les ARN de feuilles témoins et les ARN de feuilles inoculées par le mildiou, alors que la quantité d'ARN transférée semblait inférieure dans le second cas.

III.3 - Conclusions

Les résultats obtenus par Northern blot ne nous permettent pas de déterminer précisément la taille des transcrits des rétrotransposons. Aucun signal spécifique n'a été détecté pour les familles *Gentil*, *Noble* et *Tvv1*, alors que ces familles avaient été choisies car elles présentent des taux de transcription relativement forts.

Ce résultat peut s'expliquer par le fait que les sondes utilisées étaient de tailles relativement petites, respectivement 200, 259, 261 et 526 pb pour les familles *Gentil*, *Noble*, *Edel* et *Tvv1*, ce qui ne permet pas de créer des liaisons suffisamment fortes lors de l'hybridation des sondes aux transcrits. Ainsi, les sondes ont peut être été éliminées lors des lavages qui ont été réalisés avant l'étape de révélation pour éliminer l'excès de sondes non-hybridées. De plus, les sondes ciblaient spécifiquement la région *gag*, qui est considérée comme la région interne la plus variable au sein d'une famille de rétrotransposons. Les sondes ne se sont donc peut être pas hybridées correctement à tous les transcrits de la famille.

Le signal obtenu dans le cas de l'élément *Edel* est surprenant puisqu'il correspond à des transcrits de petites tailles, alors que la taille de ses ARN messagers est théoriquement supérieure à 5kb. Ce signal peut correspondre à l'hybridation de la sonde à des transcrits tronqués, à des transcrits partiellement dégradés ou à des produits de dégradation liés à un mécanisme de silencing par siARN. Cependant, la taille des transcrits correspondant à ce signal semble trop grande pour que cette dernière hypothèse soit envisagée. Quoiqu'il en soit, ces transcrits sont apparemment plus nombreux en réponse à l'induction par la blessure mécanique et par l'inoculation de mildiou, ce qui peut vouloir dire que leur présence est plus importante suite à l'induction par un stress.

Chapitre IV : Discussion de l'étude du niveau de transcription des rétrotransposons chez la vigne

Trois approches complémentaires

Le niveau de transcription des rétrotransposons du génome de la vigne a été évalué par des méthodes différentes mais complémentaires. La recherche de séquences EST dans les bases de données disponibles pour la vigne nous a permis de distinguer les familles selon leur activité transcriptionnelle. Cependant, les conditions d'obtention des banques ne sont pas toutes clairement définies et la contribution de chacune d'entre elles au nombre total de séquences EST disponibles n'est pas connue. Ces deux lacunes ne permettent pas d'évaluer avec certitude le profil d'expression et le niveau de transcription de chaque famille, mais cette analyse nous a permis d'obtenir plusieurs informations, notamment sur le nombre de familles actives au niveau transcriptionnel et sur les conditions dans lesquelles elles s'expriment.

L'évaluation du niveau de transcription par RT-PCR semi-quantitative nous a ensuite permis d'établir plus précisément le profil d'expression de 16 familles dans plusieurs tissus de la vigne, et d'évaluer l'effet de deux stress biotique et abiotique sur leur niveau de transcription.

Des résultats cohérents

Comme le résume le Tableau 12, les éléments qui n'étaient pas détectés dans les banques EST n'ont pas été amplifiés par RT-PCR sur les organes en conditions normales de croissance et n'apparaissent que rarement transcrits en condition de stress. Les familles qui étaient plus représentées dans les banques de séquences EST ont été détectées par RT-PCR. Les expériences de RT-PCR ont été réalisées sur la variété Riesling mais il est probable que l'activité de ces éléments soit comparable dans les autres cépages de l'espèce *V. vinifera*. Si trois familles ne sont apparemment transcrites dans aucune condition, neuf autres le sont dans plusieurs organes et 13 familles sont surexprimés en réponse aux stress. Comme c'est le cas chez d'autres plantes, le niveau de transcription des rétrotransposons de la vigne est, dans son ensemble, sensible à l'induction par la blessure mécanique, l'attaque de pathogène et

Famille	Détection de séquences EST	Détection par RT-PCR	
		Organes	Stress
<i>Brand</i>	-	-	-
<i>Cremant</i>	+	-	-
<i>Edel</i>	++	+	+++
<i>Gans</i>	++	+	+
<i>Gentil</i>	+	+	+++
<i>Gret1</i>	+++	+	+
<i>Huben</i>	++	+	+++
<i>Kastel</i>	+	?	?
<i>Noble</i>	+++	+++	+++
<i>Rangen</i>	+	-	+
<i>Tvv1</i>	++	+	+++
<i>Vine1</i>	+	+	+
<i>Wintz</i>	-	?	?
<i>V1</i>	+++	+	+
<i>V11</i>	+	-	-
<i>V12</i>	+	-	+
<i>V13</i>	-	-	+
<i>V14</i>	+++	-	+

Tableau 12 : Bilan de l'étude de l'activité transcriptionnelle des rétrotransposons du génome de la vigne.

probablement les changements génomiques et les dommages provoqués par la culture de tissus.

Des familles actives dans le génome de la vigne

Les familles de rétrotransposons qui possèdent un nombre de copies important dans le génome de la vigne sont également celles qui présentent une activité transcriptionnelle plus forte. Par exemple, les familles *Noble*, *Edel* et *Gentil* étaient parmi les plus représentées dans le génome. La comparaison des LTR 5' et 3' de copies appartenant à ces familles a démontré que plusieurs d'entre elles s'étaient transposées récemment dans le génome de la vigne. Ces copies présentaient également des cadres de lecture ouverts complets et sans codon stop, ce qui laisse supposer qu'elles codent pour des polyprotéines fonctionnelles. Ces indications permettent de penser que ces éléments sont capables de réaliser leur cycle de rétrotransposition de manière autonome. Cette étude pourrait être complétée par des analyses complémentaires. Comme nous l'avons vu précédemment, l'annotation des séquences EST pourra être sensiblement améliorée en considérant différemment les EST chimères. Une analyse par RT-PCR quantitative permettrait également de confirmer et de préciser les niveaux de transcription spécifiques à chaque condition.

Mobilisation des rétrotransposons

Le stress est un facteur important dans le changement de l'activité transcriptionnelle des gènes et des rétrotransposons. La transcription de gènes adjacents à des LTRs en condition de stress a été démontrée chez le blé allotétraploïde (Kashkush *et al.*, 2003), mais également chez les mammifères et la levure (Madlung and Comai, 2004). Les éléments transposables ont un effet sur le génome, même si leur activité transcriptionnelle n'est pas toujours suivie de l'insertion de nouvelles copies.

La relation entre le niveau de transcription et le taux de transposition n'a pas été clairement établie. Seules des preuves indirectes de l'augmentation de la mobilité des éléments transposables en réponse à un changement des conditions environnementales ont pu être obtenues. Chez la levure, une seule copie d'ADN est synthétisée en moyenne pour 14000 transcrits *Ty1* car des protéines de l'hôte inhibent les étapes post-transcriptionnelles de la rétrotransposition (Curcio and Garfinkel, 1999). Dans ce contexte, l'insertion de nouvelles copies dans le génome est relativement rare. L'induction de la transcription par les stress peut théoriquement augmenter les chances d'observer de nouvelles insertions.

La transposition de l'élément *mPing* chez le riz semble s'être préférentiellement produite dans des variétés provenant de régions chaudes et qui se sont adaptées au climat plus froid dans lesquelles elles sont cultivées actuellement (Jiang *et al.*, 2003b). D'après les auteurs de cette étude, il s'agit d'un exemple de stress génomique induit par le changement de température ayant participé à la diversification et à l'évolution des variétés de riz. Un élément *mPing* s'est effectivement inséré dans un gène homologue aux gènes de contrôle de la date de floraison. L'activation du rétrotransposon *BARE-1* en réponse à des changements climatiques a également été démontrée (Kalendar *et al.*, 2000). Ces observations illustrent parfaitement l'effet que peut avoir l'augmentation de la transposition sur l'adaptation d'un génome en réponse au stress provoqué par un changement des conditions environnementales.

Evaluation de la mobilité des rétrotransposons

Pour évaluer la mobilité des rétrotransposons, il est possible de rechercher l'apparition de nouvelles copies dans le génome, c'est-à-dire de détecter de nouvelles insertions. La S-SAP permet de comparer les profils d'insertion de différents génomes en amplifiant spécifiquement les régions adjacentes à chaque copie de la famille. Cette méthode est efficace pour comparer des génomes qui ont eu le temps d'évoluer séparément et d'acquérir suffisamment de nouvelles insertions pour qu'il soit possible de les distinguer. Cependant, l'insertion d'un élément transposable est un événement ponctuel dans le temps et dans l'espace, et qui n'a lieu que dans une seule cellule.

Pour qu'une nouvelle insertion soit détectable par S-SAP, il faut que la nouvelle copie soit transmise au reste de l'organisme. Chez la vigne cultivée, qui est presque exclusivement propagée par multiplication végétative, la nouvelle insertion a peu de chance de se retrouver dans la descendance de la plante. Elle peut potentiellement se propager lorsque la nouvelle insertion se produit dans la cellule d'un méristème qui se développera pour donner un nouveau rameau, mais ce cas est probablement rare et sa détection par S-SAP est peu probable. Chez la vigne, une nouvelle insertion ne peut être transmise par multiplication végétative que si elle a eu lieu dans une cellule qui sera prélevée lors du bouturage de la plante, comme dans le cas des chimères, chez lesquelles une mutation peut affecter une assise cellulaire mais pas les autres. Ainsi, l'évaluation de la mobilité des rétrotransposons de la vigne par la détection de l'apparition de nouvelles insertions est possible mais très difficile puisque la diffusion d'une nouvelle insertion est potentiellement très rare.

Participation des rétrotransposons à l'évolution du génome de la vigne

Les stress ont des effets négatifs sur la croissance et le développement des plantes puisqu'ils exercent sur l'organisme une force évolutive très importante face à laquelle l'organisme doit s'adapter et évoluer pour survivre. La mobilisation des éléments transposables, en réponse au changement des conditions environnementales, offre au génome qui les héberge un potentiel mutagène quasi-inépuisable grâce auquel l'évolution peut satisfaire son besoin perpétuel de variabilité génétique. Il est possible qu'en l'absence de toute source de variabilité génétique liée à la reproduction par voie sexuée, les rétrotransposons participent à l'évolution du génome de la vigne en lui fournissant le potentiel mutagène qui permet son adaptation à certaines conditions nouvelles.

Partie IV : Evolution de la famille

Tvv1

Chapitre I : *Etude de la variabilité structurale* *des éléments Tvv1*

La famille *Tvv1* a initialement été reconstituée par marche sur le chromosome (Pelsy and Merdinoglu, 2002). Alors que les premières données du séquençage du génome de la vigne n'étaient pas encore disponibles, la diversité structurale des éléments de cette famille a été évaluée par l'amplification PCR de sa région interne. La région couvrant la totalité de la séquence codante des éléments *Tvv1* a été amplifiée à partir d'ADN de Cabernet sauvignon. En plus des bandes de taille attendues (environ 4kb), les profils obtenus présentaient deux bandes de tailles particulièrement petites. Le séquençage et l'analyse de ces deux fragments ont montré qu'ils correspondaient à deux éléments *Tvv1* présentant de grandes délétions, *Tvv1-Δ3001* et *Tvv1-Δ3460*. Le point de délétion est caractérisé par un motif de 13 pb chez *Tvv1-Δ3001* et de 11 pb chez *Tvv1-Δ3460*. Ces deux motifs sont dupliqués dans la séquence de l'élément *Tvv1* canonique complet, et encadrent les régions délétées. Compte tenu de leur structure, ces deux éléments délétés résultent probablement d'un événement de recombinaison illégitime qui serait survenu pendant le cycle de rétrotransposition de copies entières et autonomes.

Les insertions de *Δ3001* et de *Δ3460* dans le génome de la vigne ont été estimées à 1 et 2,4 millions d'années respectivement. La répartition de ces deux éléments au sein des *Vitaceae*, parmi 58 variétés et vignes sauvages, prouve qu'ils sont présents chacun à un seul locus dans le génome et qu'ils ont été vraisemblablement distribués par croisements.

La première copie complète de la famille *Tvv1* a été identifiée dans un clone BAC de Vidal blanc. Cette copie, *Tvv1-VB*, présente toutes les caractéristiques d'une copie autonome puisque ses LTRs sont identiques et qu'elle code potentiellement pour une polyprotéine contenant tous les domaines fonctionnels nécessaires au cycle de rétrotransposition. La séquence de *Tvv1-VB* a été utilisée pour rechercher les copies de la famille *Tvv1* dans le génome de PN40024. Nous avons pu caractériser 22 copies complètes et 4 copies délétées, et étudier leur diversité.

Les résultats de cette étude sont présentés sous la forme d'un article paru dans la revue « Theoretical Applied in Genetics », disponible en ligne depuis le 10 janvier 2008.

Structural variability of *Tvv1* grapevine retrotransposons can be caused by illegitimate recombination

Cédric Moisy · Sophie Blanc · Didier Merdinoglu ·
Frédérique Pelsy

Received: 20 June 2007 / Accepted: 13 December 2007 / Published online: 10 January 2008
© Springer-Verlag 2008

Abstract Structural variability of *Tvv1*, a grapevine retrotransposon Ty1 *copla*-like family, was investigated within the grape genome and the canonical sequence of *Tvv1* determined. Then, two remarkable elements, *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640, which had suffered large deletions 3,001 bp and 3,460 bp in length of their coding sequences were compared to the canonical copy. In both deleted elements, the deletion breakpoint was characterized by a stretch 13 bp-long in *Tvv1*- Δ 3001 and 11 bp-long in *Tvv1*- Δ 3640 found duplicated in the canonical copy at each bound of the deleted regions. *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640 were both shown to be unique copies fixed at a single locus in the grapevine genome. Their presence was very variable in a set of 58 varieties and wild vines. These elements have most likely been dispersed through natural intermixing after their initial insertion whose chronology was estimated. The model that we propose to explain the structure of *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640, implies illegitimate recombination involving template switching between two RNA molecules co-packaged in the VLP prior to the integration of the deleted daughter copy into the host genome.

Introduction

Retrotransposons are mobile elements closely related to retroviruses in their structure and life cycle (Boeke and Corces 1989). Active retrotransposons transpose via the reverse transcription of a transcribed, intermediate RNA, thus increasing their copy number (Feschotte et al. 2002). The copy-and-paste cycle in the host genome comprises the synthesis of specific proteins encoded by two major genes, *gag* and *pol*. The two major classes of retrotransposons, the Ty1 *copla* (Flavell et al. 1992; Voytas et al. 1992) and the Ty3 *gypsy* (Suoniemi et al. 1998), are flanked by long terminal repeats (LTRs), but differ from each other in the order of domains inside the *pol* gene as well as in their degree of sequence conservation. The *pol* gene of Ty1 *copla*-like elements encodes from 5' the protease, integrase, reverse transcriptase and RnaseH domains. These enzymatic functions are required to synthesize an extra-chromosomal DNA daughter copy from the intermediate RNA prior to its reinsertion into the genome. The *gag* gene encodes proteins that assemble into virus-like-particles (VLP) in which the genomic RNA is packaged along with the associated enzymes and primer tRNA. Ty1 and Ty3 retrotransposons VLPs contain a dimeric RNA genome similar to what has been described for retroviruses (Feng et al. 2000; Nymark-McMahon et al. 2002). However, while in retroviruses the dimer consists of two identical plus-strand RNAs joined by non-covalent bonds between stem-loops on the monomers, the mechanism of union between monomers is yet unknown for retrotransposons, although it is conceivable that they are joined to each other through the primer tRNA_i^{Met} molecules (Feng et al. 2000).

The reverse transcription mechanism is highly error-prone due to the lack of proofreading repair activity of reverse transcriptase and RNA polymerase.

Communicated by C. Gebhardt.

C. Moisy · S. Blanc · D. Merdinoglu · F. Pelsy (✉)
UMR 1131 Santé de la Vigne et Qualité du Vin, INRA,
Université Louis Pasteur de Strasbourg,
68000 Colmar, France
e-mail: pelsy@colmar.inra.fr

Misincorporation or insertion–deletions result in base substitutions, premature stop codons or frameshifts in the new DNA copies giving rise to divergent sequences from the canonical structure (Preston 1996). Therefore, a hallmark of retroelements is the genetic heterogeneity of individual copies belonging to a same family (Casacuberta et al. 1995). Retrotransposons are ubiquitous in the plant kingdom and structural variability within different families of plant retrotransposons has been described extensively. Several reports indicate that despite the functional importance of LTR sequences, they are one of the most rapidly evolving retrotransposon regions (Casacuberta and Grandbastien 1993; Casacuberta et al. 1995; Vicient et al. 2005). Variability within retrotransposon families also includes elements lacking coding capacity for mobility-related protein, among them solo-LTRs or tandem arrays of LTRs and internal domains resulting from intrachromosomal recombination between LTRs of the same or different inserted elements (Shirasu et al. 2000; Vicient et al. 1999). Other groups of non-autonomous LTR retrotransposons comprising two LTR flanking a non-functional central domain have also been identified in plant genomes: TRIMs (terminal-repeat retrotransposons in miniature) are short LTR elements (500 bp maximum) present in both dicotyledons and monocotyledons; *Morgane* is a 1,871 bp-long element related to the *gypsy* class from wheat that displays an internal non-functional *pol*-like domain (Sabot et al. 2006), and finally LARDs (large retrotransposons derivatives) derived from *gypsy* elements have been described in barley and rice (Kalendar et al. 2004). Although non-autonomous, TRIMs, *Morgane* and LARDs are multi-copy elements that are proposed to be *trans*-mobilized by the protein complex of intact retrotransposons (Kalendar et al. 2004). Chimerical elements resulting from recombination events between close retrotransposons families can also be observed: for example, *BARE-2* is a chimerical element related to the *BARE-1-Wis-2* retrotransposon grass family produced by events of recombination between *BARE-1* and *Wis-2a*. (Vicient 2005). Intrachromosomal recombination is not the only mechanism at the origin of non-canonical elements: template switching between the two packaged RNA molecules during retrotransposition has also been proposed as a mechanism leading to deleted copies. This mechanism has been put forward to explain the structure of *wsStonor*, an element belonging to the *Stonor* maize retrotransposon family that had suffered a 2.6 Kb-long deletion of the entire *gag* region and part of the protease domain (Marillonnet and Wessler 1998).

The grapevine genome sequence, sized 487 Mb, has recently been obtained from the highly homozygous genotype PN40024. The current annotation of the transposable elements revealed a large prevalence of retrotransposons over transposons, most of which (47%) correspond to *Ty1*-

copia like retrotransposons. Distribution of transposable elements over the grape chromosomes is mainly restricted to low gene-density regions and introns (The French–Italian Public Consortium for Grapevine Genome Characterization 2007). Before the grape genome sequencing, three grapevine retrotransposons were described: a unique copy of *Vine-1*, a *Ty1-copia* like element (Verriès et al. 2000) and of *Gret 1*, a *Ty3-gypsy* like element (Kobayashi et al. 2004) and the *Tvv1* family whose consensus sequence was reconstituted by chromosome walking (Pelsy and Merdinoglu 2002). It was shown that in a single genome, *Tvv1* elements shared a single, highly conserved 4,143 nucleotides long ORF and possessed LTRs sized between 149 and 157 bp and an untranslated leader (UTL) region, upstream of the ORF, highly variable in size and sequence (Pelsy and Merdinoglu 2002). In the present study, we have investigated the structural variability of the *Tvv1* family in the grapevine genome to determine the canonical sequence. Then, we have analyzed the structure and dispersion of two remarkable *Tvv1* elements that had suffered large deletions of their coding sequence. Based on their structural and genetic characteristics, we found strong evidence supporting our hypothesis of illegitimate recombination during reverse transcription to explain the underlying mechanism leading to the formation of these deleted elements.

Materials and methods

Plant material and DNA extraction

The two deleted *Tvv1* elements were characterized in cabernet sauvignon (certified clone 169; Anonymous 1995). Self-progenies were derived from cabernet sauvignon (Chilean clone), muscat Ottonel (certified clone 59), pinot noir (certified clone 162) and riesling (certified clone 49). Distribution of the deleted elements was evaluated in a collection of 58 *V. vinifera* accessions, represented by wine and table grapevine (52 varieties), and wild vines (6 accessions) belonging to the ampelographic collection of INRA-Colmar (France).

Young, expanded leaves were collected from individual plants and ground into fine powder with liquid nitrogen. Total genomic DNA was purified using Dneasy™ Plant Mini-Kit (Qiagen, Hilden, Germany) as described by the supplier.

PCR conditions and fragment analysis

All primers were synthesized by MWG Biotech AG (Ebersberg, Germany); their sequences and location are given in Table 1 and Fig. 1. Long-range amplifications

Table 1 Primers used in the study

No.	Primer	Orientation	Sequence (5'–3')	Location	Position/ <i>Tvv1</i> -VB
1	P18	Sense	GTTGCTGATATAGTGCCTATTGTGTCA	ORF	939–965
2	Pltr4	Antisense	CAATTGAGAGTCCT(AG)AATTAGGAA GGT	LTR	56–82//5129–5155
3	Pltr3	Sense	CCTTCCTAATT(CT)AGGACTCTCAATTGT	LTR	57–83//5130–5156
4	Pdel1	Antisense	TT(CT)TGCACTC(AG)AACGGGATACAACACT	ORF	4701–4727
	Pvv78xfw	Sense	AAAGGTGCAGCCAACCAAACCTCTCTTC	5' Δ 3460 host region	–
	GW16	Antisense	GTGGTCATCTTTAGCAACACTTCTCAA	ORF	1032–1058
	GW13	Antisense	GGACGATGAAAGGCATTCCATACATCA	3460bp-deletion	1250–1276
	Pvv78xrv	Antisense	GCAAGTAGTAAGGCAGGASTTAAGAGG	3' Δ 3460 host region	–
	Pdel4	Sense	GTAAGAAGCAAAGTGTGTATCCCGTT	ORF	4690–4716
	GW1	Sense	CATTGGCATGAACAAGAGCGAGAAAG	3460bp-deletion	4027–4053
	Pbo5	Sens	TGTCTATAGGCAAATGGAAGGGAECTA	5' Δ 3001 host region	–
	Pbo3	Antisense	TGGGGTGGAAACGAACGAGTGGAAATA	3' Δ 3001 host region	–
	PVB1	Sense	TGTTTCTCAT GCACATTACA AG	5' VB host region	–132 to –110
	Pltr2	Antisense	ATCCCTAAGGAAACAACCAA	LTR	33–52//5106–5125
	PVB2	antisense	CCTTGTGAGGGATAACATAACC	3' VB host region	+424 to +446
	Pltr1	Sense	CCTAATTCAGGACTCTCAAT	LTR	61–80//5134–5153

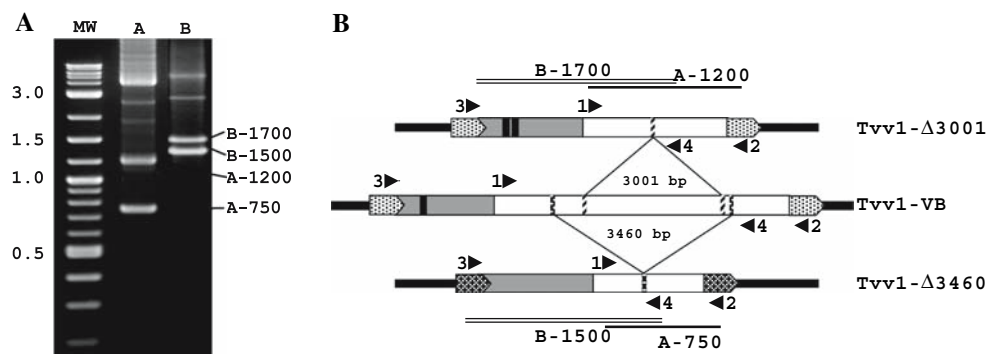


Fig. 1 Structure of the *Tvv1*-VB, *Tvv1*- Δ 3001 and *Tvv1*- Δ 3460 loci. **a** A and B PCR-amplifications of *Tvv1* elements of cabernet sauvignon. Molecular weight standard range from 0.2 to 10 Kb. Additional low intensity PCR products should result from amplification of additional deleted copies. **b** Schematic representations of the

full-length *Tvv1*-VB of Vidal blanc and of the two deleted elements *Tvv1*- Δ 3001 and *Tvv1*- Δ 3460. Arrow boxes represent LTRs, gray boxes UTL and open boxes ORF regions. Black arrows represent primers used in the study, whose numbers are reported in Table 1

were performed using the primer pair Pltr4 and P18 for A amplifications and the primer pair Pltr3 and Pdel1 for B amplifications. Amplification reactions were performed in a mixture containing 20 ng DNA, 500 nM each primer, 1.5 mM MgCl₂, 200 μ M deoxynucleoside triphosphate and 0.5 U DyNAzyme EXT™ Taq DNA polymerase (Finnzymes, Espoo, Finland) in a 20- μ l reaction buffer (DyNAzyme EXT Mg-free buffer, Finnzymes). PCR cycling conditions were 94°C for 2 min, 35 cycles of 94°C for 10 s, 55°C for 30 s and 68°C for 8 min and ended by a final extension at 68°C for 7 min. *Tvv1*-VB insertion site was PCR-amplified from primers designed in the 5' (PVB1) or 3' (PVB2) flanking regions of *Tvv1*-VB and in the LTR, Pltr1 and Pltr2, respectively. *Tvv1*- Δ 3001 flanking regions

were identified by a genome walking procedure on a BAC clone consisted in two successive PCR amplifications (Pelsy and Merdinoglu 2002), the 5' region was amplified using primers in antisense orientation GW16 and Pltr4 and the 3' region was amplified using primers in sense orientation Pdel4 and Pltr3. *Tvv1*- Δ 3460 flanking regions were amplified using primers Pvv78x designed from the pinot noir sequence upstream and downstream from the *Tvv1*- Δ 3460 element, Pvv78xfw used in pair with primer GW16 and Pvv78xrv with Pdel4. Amplification reactions were performed in a mixture containing 10 ng DNA, 500 nM each primer, 150 μ M deoxynucleoside triphosphate and 0.5 U GoTaq® DNA polymerase (Promega Corporation, Madison, WI, USA) in a 20- μ l GoTaq® reaction buffer.

PCR cycling consisted of 7 cycles of 94°C for 10 s, and 70°C for 3 min, followed by 32 cycles of 94°C for 10 s, and 65°C for 3 min followed by an elongation step of 65°C for 10 min. All amplifications were performed on a GeneAmp® PCR System 9700 thermocycler (PE Applied Biosystems, Foster City, CA, USA). Aliquot of the PCR products were analyzed by electrophoresis on a 2% agarose gel in 1 × TAE (40 mM TRIS-acetate pH 7.8, 20 mM EDTA) loading 2-Log DNA ladder as a standard of the molecular weight (New England BioLabs Inc., Ipswich, MA, USA).

For bands of interest, PCR fragments were cloned into the pCR® 4-TOPO cloning vector (Invitrogen, Groningen, NL, USA), according to the supplier instructions. Two random clones were selected for forward and reverse sequencing by PCR using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Samples were processed on an ABI3100 Genetic analyzer, a 16-capillary electrophoresis sequencing instrument. Computer-assisted analysis of the sequence data was performed using DNAsis 2.1 (Hitachi Software Engineering Co, Ltd) and Vector NTI 7 (InforMax Inc.). The nucleotide data of the 12 reported fragments have been deposited in GenBank database under accession numbers EU293346-EU293357. *Tvv1* elements were searched in the grapevine genome using Genoscope BLAST server (http://www.cns.fr/cgi-bin/blast_server/projet_ML/blast.pl). They were individually examined for position of their LTRs, UTL, ORFs using Vector NTI 7 (InforMax Inc.).

Results

The *Tvv1* family

The consensus sequence of *Tvv1* grapevine retrotransposons was first reconstituted by chromosome walking from the variety riesling (Pelsy and Merdinoglu 2002). A sequence homologous to an expected *Tvv1* element was identified in the BAC clone A1-16 (GenBank accession no. EU304807) from the grapevine variety *Vidal blanc* (Zyprian et al. personal communication). This clone contained a full-length *Tvv1* copy sized 5,222 bp, referenced as *Tvv1*-VB, whose structural characteristics are two perfect 149-bp long LTRs flanking an internal domain. *Tvv1*-VB LTR sequence is identical to *Tvv1* LTR L3B3 previously identified by chromosome walking (GenBank accession no. AF478367), but while the *Tvv1*-VB is flanked by perfect 5-bp direct repeats (ATTCT), the 5-bp downstream of the LTR L3B3 (TTTTC) indicate that the two insertion sites are different. The 4,924 bp-long internal domain includes the priming motifs purine binding site (PBS) and polypurine tract (PPT) described for all *Tvv1* elements and encode

a putative 1,382 amino acid-long polyprotein within a single translational reading frame, uninterrupted neither by stop-codons nor by frameshifts, showing 98.9% identity to the *Tvv1* consensus sequence. UTL region, starting by the PBS complementary to the tRNA^{met}, is 768 bp-long, a size in the range of expected UTL found in the riesling variety (Pelsy and Merdinoglu 2002). No tri-nucleotide ATG that could indicate the starting point of a putative open reading frame was found in the UTL region. The GC content of *Tvv1*-VB is 42.5%, but it varies along the retrotransposon sequence: from 30% at the LTRs, to 41.5% at the ORF and to 53.5% at the UTL. The presence of *Tvv1*-VB at the predicted locus in the genome of *Vidal blanc* was assessed in the variety cabernet-sauvignon by a PCR-amplification using primers designed in the 5' and 3' flanking regions of the element and in the highly homozygous genotype PN40024 by a bioinformatic analysis. Only the empty site was amplified in cabernet sauvignon and found in the sequence of PN40024 (ML8X_scaffold_43).

Later on, 26 *Tvv1* elements flanked by 5-bp direct repeats were identified in the PN40024 sequence (Table 2). A total of 22 copies were full length; their sizes ranged from 4,640 to 7,162 bp, with LTRs from 149 to 198 bp, UTL from 179 to 1,069 bp and *gag-pol* gene from 4,139 to 6,114 bp. Four of them displayed identical LTRs and an uninterrupted single ORF, indicating that these copies were both recent and potentially functional. None of them were fully identical to *Tvv1*-VB, the only PN25 displayed a UTL region 99% similar to the *Tvv1*-VB, but differed from *Tvv1*-VB by a 8 bp deletion in the *gag-pol* region leading to a non-functional copy. Taking into consideration elements that display identical LTRs and duplication sites, it could be considered that the canonical *Tvv1* element has 149bp-long LTRs, a 4,146bp-long ORF encoding a putative 1,382 amino acid-long polyprotein within a single translational open reading frame. PN120 and PN139 have these characteristics, but differ in their UTL region sized 824 and 891 bp, and show 69.6% identity. In addition to full-length copies, 13 incomplete copies were characterized and 4 others suffering deletions, among them PN16 will be extensively described later. Twenty *Tvv1* elements were located on independent scaffolds and three scaffolds contained two elements. Identified as a canonical element, *Tvv1*-VB sequence has been considered as the reference in the following analysis.

Two deleted copies of *Tvv1* from cabernet sauvignon

To amplify *Tvv1* copies, PCR amplifications were performed using primers P18, located at the very beginning of the ORF, and Pltr4, in the consensus part of the LTR (A amplification; Table 1). Amplification of cabernet

Table 2 Sequence characteristics of *TvvI*-VB and 26 *TvvI* elements identified in the grapevine genome sequence

<i>TvvI</i> number	Copy state	Total length (bp)	Direct repeat 5'	LTR		UTL (bp)	gag-pol gene (bp)	Direct repeat 3'	Location in PN40024	
				(bp)	Identity (%)				Chromosome	Scaffold
VB	Full	5,222	ATTCT	149	100.0	768	4146	ATTCT	10	scaffold_43 ^a
PN120	Full	5,278	GCCAT	149	100.0	824	4146	GCCAT	18	scaffold__1
PN139	Full	5,345	TTCAT	149	100.0	891	4146	TTCAT	1	scaffold_136
PN121	Full	5,525	AAGAA	150	100.0	1,069	4146	AAGAA	6	scaffold_3
PN140	Full	5,226	AGGATA	149	100.0	743	4152	AGGATA	6	scaffold_28
PN25	Full	5,215	CCTTT	149	100.0	768	4139 ^b	CCTTT	8	scaffold_115
PN136	Full	5,075	GGCCC	151	99.3	620	4143 ^b	GGCCA	12	scaffold_93
PN142	Full	5,256	AGAAC	149/150	99.3	803	4144 ^b	AGAAC	5	scaffold_156
PN124	Full	5,268	TTATT	151	99.3	813	4143 ^b	TTATT	14	scaffold_54
PN123	Full	6,387	GTTAC	198	99.3	511	5470 ^b	GTTAC	7	scaffold_44
PN145	Full	7,162	GCCTT	149	99.3	740	6114 ^b	ACCTT	14	scaffold_63
PN135	Full	5,351	AATGG	150	98.7	898	4143 ^b	AATGG	18	scaffold_61
PN24	Full	5,532	CAATT	151	98.7	1,076	4144 ^b	CAATT	14	scaffold_54
PN141	Full	5,307	TGTTT	150	98.0	843	4154 ^b	TGTTG	7	scaffold_20
PN128	Full	5,362	AAATG	169	97.6	858	4156 ^b	ATATG	nc	scaffold_754
PN129	Full	5,522	GAACA	168	97.0	1,032	4144 ^b	GAACA	nc	scaffold_109
PN131	Full	4,958	CATTA	179	96.6	466	4124 ^b	CATTC	10	scaffold_179
PN134	Full	4,640	AGAAG	167/168	95.2	179	4116 ^b	GGAAG	nc	scaffold_139
PN04	Full	5,027	AACAA	172/168	93.1	511	4166 ^b	AACAA	5	scaffold_2
PN137	Full	4,666	AGATG	182/194	90.7	417	3863 ^b	AGATG	2	scaffold_140
PN126	Full	4,910	ATGTC	194/182	90.7	408	4116 ^b	ATGTC	6	scaffold_28
PN01	Full	5,498	ACTCC	186/168	86.6	983	4152 ^b	ACTCC	nc	scaffold_80
PN122	Full	5,651	ATTAT	161/189	79.4	960	4331 ^b	ATTAT	16	scaffold_10
PN132	Deleted	5,116	GAAGG	151/29 ^c	nc	782	4144 ^b	GAAGG	7	scaffold_20
PN144	Deleted	4,259	AATTA	150/156	93.6	nc	nc	AATTC	8	scaffold_88
PN02	Deleted	2,481	ATTAC	59 ^c /150	nc	nc	nc	ATTAC	18	scaffold_24
PN16	Deleted	2,057	CCCAT	161/163	95.7	1,038	685 ^b	CCCAT	8	scaffold_99

Nucleotides that are different in the direct repeats flanking an element are given in bold

^a Location of the *TvvI*-VB empty site

^b ORF interrupted by a stop codon

^c LTR partially deleted

sauvignon genomic DNA creates a major product 4.3 Kb-long, in agreement with the structural organization of the full-length *TvvI* elements, and two additional, strongly amplified, bands sized 750 bp and 1,200 bp named A-750 and A-1200, respectively (Fig. 1). Sequencing of these additional bands showed that they could belong to two different *TvvI* elements suffering an internal deletion. This result was confirmed by a second PCR amplification using primers Pltr3, designed in the most conserved region of the LTR, and Pdel1, located downstream of the deletions (B amplification, Fig. 1; Table 1). In addition to the expected bands around 4Kb-long, cabernet sauvignon showed two additional bands sized 1,500 and 1,700 bp, named B-1500 and B-1700, respectively. The four unexpected PCR products (bands A and B) were cloned and sequenced.

TvvI-Δ3001

Two cloned A-1200 fragments displayed almost identical sequences (two mismatches/1,216 bp) as well as two cloned B-1700 fragments (three mismatches/1,703 bp). Sequence analysis revealed that A-1200 and B-1700 overlapped and, when compared to the sequence of *TvvI*-VB, they displayed the same 3,001 bp-long deletion. This result led us to believe that these two bands originated from the same deleted copy that we called *TvvI*-Δ3001. A cabernet sauvignon BAC library was screened to find a clone displaying the deleted element (Lamoureux et al. 2006). The 5' and 3' regions overlapping *TvvI*-Δ3001 LTRs and host flanking regions were determined by genome walking on the BAC clone 3E11. *TvvI*-Δ3001 is sized

2,256 bp-long and is flanked by identical 5-bp direct repeats (TGTAC). The two *TvvI*- Δ 3001 LTRs, 150 bp-long, differ by two nucleotides and show 96% identity with *TvvI*-VB LTRs (four substitutions and one single nucleotide indel; Fig. 2). Downstream of the 5'LTR, the UTL 801-bp long starts with the PBS complementary to the tRNA^{met}. This UTL sequence is nearly identical to the *TvvI*-VB UTL except for the addition of 33 bp. This additional stretch, duplicated in direct orientation in the *TvvI*- Δ 3001, is already present in the *TvvI*-VB UTL. In *TvvI*- Δ 3001 UTL, the 33 bp-repeats are separated by 10 nucleotides (Fig. 2). The deletion breakpoint of *TvvI*-3001 located in position 1,649–4,650 in *TvvI*-VB is characterized by the presence of a 13 bp-long stretch. In the full-length *TvvI*-VB, two nearly identical repeats in direct orientation flank the sequence corresponding to the deletion: the 3' stretch is identical to the deletion breakpoint of *TvvI*- Δ 3001, while the 5' stretch differs from the 3' stretch by one mismatch (Table 3). Hypothetical translation of the remaining ORF sequences upstream of the deletion started with the regular ATG and revealed a single putative sequence of 252 amino acids in the *gag* region uninterrupted by a stop-codon (Fig. 3). The first 247 amino acids differed by 4 amino acids from the corresponding *TvvI*-VB

protein, and the deletion led to the addition of 5 amino acids followed by a stop codon. Regarding the sequence downstream of the deletion, the hypothetical translation revealed a regular sequence of 134 amino acids that differs by 3 amino acids from the corresponding *TvvI*-VB region (Fig. 3).

Because of the mechanism of transposition, the two LTRs of a retrotransposon are theoretically identical at the time of their insertion (Jordan and McDonald 1998, 1999; SanMiguel et al. 1998). Thus, calculation of the base substitution rate between the two *TvvI*- Δ 3001 LTRs, 150 bp-long and differing by two mismatches, was used to evaluate the timing of insertion. We used the average base substitution rate derived from the grass *adh1-adh2* region of 6.5×10^{-9} substitution per site per year (Gaut et al. 1996), given the assumption of a uniform mutation rate, to estimate that *TvvI*- Δ 3001 inserted into its present location ≈ 1 million years ago.

TvvI- Δ 3460

As for *TvvI*- Δ 3001, sequence analysis of A-750 and B-1500 from cabernet sauvignon leads to the structure of

	A	B
LTR- Δ 3001	1 TGTAGCTGT ATATATCTGT ACATACCAGA ATTAAGTTGT TTCCTTAGGG ATAATACCTT CCTAATTCAG GACTCTCAAT TGTATATATA AACAAATATT	
LTR-VB	1 TGTAGCTGT ATATATCTGT ACATACCATA ATTTGGTTGT TTCCTTAGGG ATAATACCTT CCTAATTCAG GACTCTCAAT TGTATATATA AACAAATTAT	
LTR- Δ 3001	101 ATTCATCTAA TAAAAGACAA GGAATTGAGA ATTACCTTGA TTCGGTTACA	
LTR-VB	101 ATTCATCTAA TAAA-GACAA GGAATTGAGA ATTACCTTGA TTCGGTTACA	
	PBS *****	
UTL- Δ 3001	151 TGGTATCAGA GCCACGTTTG GTTCAAGGGA ATTTTCTGGG TTTTGGGTTT TGTTCCTAAC AATCTGGAAY AGTAACCTGT TACGTGTTTG GTCTGGATTG	
UTL-VB	149 TGGTATCAGA GCCACGTTTG GTTCAAGGGA ATTTTCTGGG TTTTGGGTTT TGTTCCTAAC AATCTGGAAC AGTAACCTGT TACGTGTTTG GTCTGGATTG	
UTL- Δ 3001	251 CGTCACCGCC GACCCAGGGA GGAGCGCATT CTTGTTTCCG GAGCGTGGGA TGCTTGAGG CCGCCGTTTC TTCACCFATT TGACGTCCTCG ACGTTGCTTG	
UTL-VB	250 CGTCACCGCC GACCCAGGGA GGAGCGCATT CTTGTTTCCG GAGCGTGGGA TGCTTGAGG CCGCCGTTTC TTCACCGATT TGACGTCCTCG ACGTT-----	
UTL- Δ 3001	351 AGGCCCGCGT TTCTTACCAG ATTTGACGTC CCATCTCTCG ATCTGACTTT CCGACGTGTT CGATCAACTG TCGAGGGATC GCAAAGATTC ATCGACGACT	
UTL-VB	345 -----TC -----TC CCATCTCTCG ATCTGACTTT TCGACGTGTT CGATCAACTG TCGAGGGATC GCAAAGATTC ATCGACGACT	
UTL- Δ 3001	451 GTGGAATCGC TGTGCGCTTC TTCTCAGATC CCGCCACCCT CTCACGCCTT CTCTTCGCCC CGTTCCTCGT CAACGCGGTC TTCACCTCGG CCGGAACCCT	
UTL-VB	417 GTGGAATCGC TGTGCGCTTC TTCTCAGATC CCGCCACCCT CTCACGCCTT CTCTTCGCCC CGTTCCTCGT CAACGCGGTC TTCACCTCGG CCGGAACCCT	
UTL- Δ 3001	551 AAGCCCCCA CCGAAAACCT AGCACGACGC TTTTCTCAT CCGCAGCACC GTGTTTGCA CCAAATCCAC GCCAGAAATC AAACAGTCGG CGTCGTCGGA	
UTL-VB	517 AAGCCCCCA CCGAAAACCT AGCACGACGC CTTTCTCAT CCGCAGCACC GTGTTTGCGC TCAAATCCAC GCCAGAAATC AAACAGTCGG CGTCGTCGGA	
UTL- Δ 3001	651 GATCATTGAA AGCCCCGGCC TCCATTCTTC TCGGGCCCGA CACACGACCC ACGTTTGCA GCAGCCCTGC TCTGTCTCCG GTCGTGCTG CCGGACCGAT	
UTL-VB	617 GATCATTGAA AGCCCCGGCC TCCATTCTTC TCGGGCCCGA CACACGACCC ACGTTTGCA GCAGCCCTGC TCTGTCTCCG GTCGTGCTG CCGGACCGAT	
UTL- Δ 3001	751 CTGAGTTTCT TCGCCCGCCT TGGTCGCCAC GACTCAGCAG TTTCAGGGTG AGAAAGAGAG GGAGTCAAGG GTTTGACGCA CGCTTTTACC CCTTTTACCC	
UTL-VB	717 CTGAGTTTCT TCGCCCGCCT TGGTCGCCAC GACTCAGCAG TTTCAGGGTG AGAAAGAGAG GGAGTCAAGG GTTTGACGCA CGCTTTTACC CCTTTTACCC	
UTL- Δ 3001	851 TTTTGTCCACC TGGTAAGAGA GGTATTGTGC TTTTCTTCT TGGAGTGGTA AATTGCAATT TCTGGTCTGG CTGATTGCTT TGAAGATTAA ATTTTATTGT	
UTL-VB	817 TTTTGTCCACC TGGTAAGAGA GGTATTGTGC TTTTCTTCT TGGAGTGGTA AATTGCAATT TCTGGTCTGG CTGATTGCTT TGAAGATTAA ATTTTATTGT	
UTL- Δ 3001	951 c	
UTL-VB	917 c	

Fig. 2 LTR and UTL sequences of *TvvI*- Δ 3001 from cabernet sauvignon. Nucleotides that are different in *TvvI*- Δ 3001 and *TvvI*-VB are underlined. Solid line (PCR A) and double solid line (PCR B)

overlie the parts of *TvvI*- Δ 3001 revealed by PCR-amplification. The sequence duplicated in *TvvI*- Δ 3001 UTL is in box. Dotted line overlies PBS

Table 3 Size, deletion and deletion breakpoints of *TvvI*- Δ 3001 and *TvvI*- Δ 3460

	<i>TvvI</i> - Δ 3001 CS	<i>TvvI</i> - Δ 3460 CS
Element size (bp)	2,256	2,074
Deletion size (bp)	3,001	3,460
Deletion location/ <i>TvvI</i> -VB	1,649–4,650	1,222–4,682
Sequence breakpoint in the deleted element	ATTGTGTGTTCTT	TGGAAAAGGAA
Homologous sequences to the	ATTGTGTGTTTTT	TGGAAAAGGAA
breakpoint in <i>TvvI</i> -VB	ATTGTGTGTTCTT	TGGAAAAGTAA

TvvI- Δ 3460, a deleted *TvvI* element that had suffered a 3,460 bp-long deletion. The full sequence of *TvvI*- Δ 3460 was also found in the sequence of pinot noir (2,074 bp; accession number AM433636.2) and of PN40024 (PN16, 2057 bp; Table 2) at the same locus, giving access to the host flanking sequences of this element. Location of *TvvI*- Δ 3460 in cabernet sauvignon was confirmed by sequencing PCR products amplified from primers (Pvv78x forward and reverse) designed in the host flanking sequences and its full size deduced (2,060 bp). In the three genomes, *TvvI*- Δ 3460 element is flanked by identical 5-bp direct repeats (CCCAT) indicating a unique integration event.

In cabernet sauvignon, *TvvI*- Δ 3460 shows 5' LTR and 3' LTR, 165 and 161 bp-long, respectively. The two LTRs differ in the number of TA motifs in a microsatellite stretch, 11 for the 5' LTR and 9 for the 3' LTR, in addition to five substitutions. It is also the difference in the number of TA motifs that justify the size of *TvvI*- Δ 3460 that is different in the three genomes. The 3'LTR of *TvvI*- Δ 3460 shows 76.8% identity with the *TvvI*-VB LTRs. Downstream of the 5'LTR, the UTL region, 1,038 bp-long, starts with the PBS complementary to the tRNA^{met}. *TvvI*- Δ 3460 UTL sequence shows 18.2% identity with *TvvI*-VB UTL, differences consisting of five insertions from 138 to 5 bp-long and by many substitutions. In *TvvI*- Δ 3460, the deletion breakpoint located in position 1,222–4,682 in *TvvI*-VB is characterized by the presence of a 11 bp-long stretch. Similar to *TvvI*- Δ 3001, two almost identical short repeats in direct orientation flank the sequence corresponding to the deletion in the full-length *TvvI*-VB: the 5' stretch is identical to the deletion breakpoint of *TvvI*- Δ 3001, while the 3' stretch differs from the 5' stretch by one mismatch (Table 3). Hypothetical translation of the ORF sequences upstream of the deletion starts with the regular ATG and reveals a single putative sequence of 113 amino acids in the *gag* region uninterrupted by stop-codon. The first 101 amino acids are nearly identical to the corresponding *TvvI*-VB protein (4 amino acids are different) and the deletion leads to the addition of 9 amino acids before being interrupted by a stop codon. Regarding the sequence downstream of the deletion, the hypothetical translation reveals a sequence of 126 amino acids that differs by 4 amino acids from the corresponding *TvvI*-VB region

(Fig. 3). Cabernet sauvignon and pinot noir *TvvI*- Δ 3460 copies mainly differed in the number of TA motifs in the LTRs and also by 2 single nucleotides in the internal domain.

In the same way as for *TvvI*- Δ 3001, the timing of insertion of *TvvI*- Δ 3460 into its present location was estimated at ≥ 2.4 millions years ago, taking into consideration the five single nucleotide polymorphism but not variation in the number of repeats of the microsatellite stretch.

Segregation analysis

To determine whether *TvvI*- Δ 3001 and *TvvI*- Δ 3460 insertions are homozygous or heterozygous, transmission of B-1500 and B-1700 bands through sexual reproduction was studied. A self-progeny of cabernet sauvignon was available but consisted of only 16 individuals. To increase the number of plants, we performed the segregation analysis on 121 plants of a self-progeny of muscat Ottonel, a variety that displayed both *TvvI*- Δ 3001 and *TvvI*- Δ 3460. We previously checked that the flanking regions of *TvvI*- Δ 3001 and *TvvI*- Δ 3460 in muscat Ottonel were identical to those of other accessions, indicating that both deleted elements were at the expected locus. In both progenies, the presence versus absence of B-1700 and B-1500 was scored as segregation markers of *TvvI*- Δ 3001 and *TvvI*- Δ 3460, respectively. Eight selfed seedlings of cabernet sauvignon and 87 of muscat Ottonel displayed B-1700, while 12 selfed seedlings of cabernet sauvignon and 80 of muscat Ottonel displayed B-1500. A segregation distortion of cabernet sauvignon self-seedlings was observed, which could have probably resulted from the low number of seedlings available. The presence of *TvvI*- Δ 3001 in 78.2% of the muscat Ottonel seedlings indicated that *TvvI*- Δ 3001 segregate was a single heterozygous Mendelian locus. In the same way, the presence of *TvvI*- Δ 3460 was shown in 75% of the cabernet sauvignon and 72.1% of the muscat Ottonel self-seedlings. We assumed that only one copy of *TvvI*- Δ 3001 and one copy of *TvvI*- Δ 3460 were present in muscat Ottonel and also in cabernet sauvignon, although five positive clones containing *TvvI*- Δ 3001 were identified

Tvv1-VB	1	MEEKNSVAD	IVPIVSKITE	HKLNGSNYIE	WNKTIKIYLR	SVAKDDHLTE	EPPNDHTRKL
Tvv1-Δ3001	1	MEEKNSVAD	IVPIVSKITE	HKLNGSNYIE	<i>W</i> S KTIKIYLR	SVAKDDHLTE	EPPNDHTRKL
Tvv1-Δ3460	1	MEEKNSVAD	IVPIVSKITE	HKLNGSNYIE	<i>W</i> S KTIKIYLR	SVAKDDHLTE	EPPND N TRKL
Tvv1-VB	61	WMQDDARLFL	QMKNSINSDI	VGLLSHCEFV	KELMDYLDFL	YSGKGNVSRM	YDVWNAFHCP
Tvv1-Δ3001	61	WMQDDARLFL	QMKNSINSDI	VGLLSHCEFV	KELMDYLDFL	YSGKGNVSRM	YDVWNAFHCP
Tvv1-Δ3460	61	WMQDDARLFL	<i>Q</i> I KNSINSDI	VGLLSHCEFV	KELMDYLDFL	<i>Y</i> F GKVRSKVLYPV*	
Tvv1-VB	121	EKGAKSLTAY	FMDFKKVYEE	LNALMPFSPD	VRVQQAQREQ	MAVMSFLSGL	PSEFETAKSQ
Tvv1-Δ3001	121	EKGAKSLTAY	FMDFKKVYEE	LNALMPFSPD	VRVQQAQREQ	MAVMSFLSGL	PSEFETAKSQ
Tvv1-Δ3460							
Tvv1-VB	181	ILSGSDIGSL	QEVFSRVLRT	ENVSSSQHTN	VLIAKRENAE	NARRVNNRGG	NRAFENRGND
Tvv1-Δ3001	181	ILSGSDIGSL	QEVFSRVLRT	ENVSSSQHTN	<i>VL</i> V AKGENAE	NARRVNNRGG	NRAFENRGND
Tvv1-Δ3460							
Tvv1-VB	241	SSTTIVCFYC	HEAGHTKKNK	RKLQNRNRRI	QTANVATSDT	ATFSDSSNKI	VTMTAEFSK
Tvv1-Δ3001	241	SSTTIVC <i>SLA</i>	<i>GI</i> *				
Tvv1-Δ3460							
Tvv1-VB					<900 aa>		
Tvv1-VB	1201	QILCYLKKAP	GLGILYSSQG	HTRIECFSDA	DWAGSKFDRR	STTGVCVFFG	GNLVAWKSKK
Tvv1-Δ3001							<i>FG</i> GNLVAWKSKK
Tvv1-Δ3460							KSKK
Tvv1-VB	1261	QSVVSRSSAE	SEYRAMSQAT	CEIIWIHQLL	CEVGMKCTMP	AKLWCDNQAA	LHIAANPVYH
Tvv1-Δ3001		QSVVSRSSAE	SEYRAMSQAT	CEIIWIHQLL	CEVGMKCTMP	AKLWCDNQAA	LHIAANPVYH
Tvv1-Δ3460		QSVVSRSSAE	<i>S</i> K YRAM A QAT	CEIIWIHQLL	CEVGMKCTMP	AKLWCDNQAA	LHI V ANPVYH
Tvv1-VB	1321	<i>ERTKHIEVDC</i>	HFIREKIEEN	LVSTGYVKTG	<i>EQLGDIPTKA</i>	LNGTRVEYFC	NKLGMINIYA
Tvv1-Δ3001		<i>ERTKHIEVDC</i>	<i>H</i> L IREKIEEN	LVSTGYVKTG	<i>EQLGDIPTKA</i>	LNGTRVEYFC	NKLGMINIYA
Tvv1-Δ3460		<i>ERTKHIEVDC</i>	HFIREKIEEN	LVSTGYVKTG	<i>EQLGDIPTKA</i>	LNGTRVEYFC	NKLGMINI Y
Tvv1-VB	1381	PA*					
Tvv1-Δ3001		PA*					
Tvv1-Δ3460		PA*					

Fig. 3 Comparison of the amino acid sequences of *Tvv1*-VB, *Tvv1*-Δ3001 and *Tvv1*-Δ3460. Amino acids different from *Tvv1*-VB are in *bold* and *underlined*. Full line box shows the nucleic acid binding domain of the *gag* protein and dotted line boxes motifs show I and II

by screening the 6X cabernet sauvignon BAC library (Lamoureux et al. 2006).

The presence of a full-length element at the *Tvv1*-Δ3001 insertion site in absence of the deleted element was investigated by amplifications using primers Pbo5 and Pbo3, designed from the host flanking regions of the cabernet sauvignon BAC clone, in 24 self-seedlings of pinot noir and of riesling, two varieties that do not display *Tvv1*-Δ3001 and in the 16 self-seedlings of cabernet sauvignon. All pinot noir and riesling self-seedlings showed a band corresponding to the expected empty site, 600 bp-long while 8 self-seedlings of cabernet sauvignon displayed a band sized 2,800 bp corresponding to the full site in agreement with the segregation of B-1700 and 12 seedlings the 600 bp band corresponding to the empty site. This result shows that the insertion site of *Tvv1*-Δ3001 is empty in absence of the deleted element. As no segregating

of the RNaseH domain. Theoretical addition of amino acids to the deleted element sequences is indicated in *italic*. For clarity, amino acids of *Tvv1*-VB absent in both deleted elements are not shown, but their number is indicated between *angle brackets*

population was available for varieties that do not displayed *Tvv1*-Δ3460, in order to reveal full-length *Tvv1* at the *Tvv1*-Δ3460 insertion site, amplifications were carried out in several varieties using primers Pvv78x designed from the host flanking regions of *Tvv1*-Δ3460 in pair with GW1 and GW13 (Table 1) designed within the 3460 bp-deletion. The results of these PCR were negative, indicating that no full-length *Tvv1* copy was present at the locus of *Tvv1*-Δ3460.

Distribution of *Tvv1*-Δ3460 and *Tvv1*-Δ3001 within a *V. vinifera* collection

Distribution of bands produced by B amplification was evaluated in a collection of 58 cultivated grapevines and wild vines. All accessions displayed B-4000 bands, most

probably indicating the presence of full-length *Tvv1* elements in their genome. Of the 58, 55 accessions amplified B-1500 and 10 accessions amplified B-1700 corresponding to *Tvv1*- Δ 3460 and *Tvv1*- Δ 3001, respectively (Table 4). Four accessions amplified neither B-1500 nor B-1700. Nevertheless, no accessions were found that amplified B-1700, but not B-1500.

Discussion

Canonical *Tvv1* element

Based on the structure of different full-length copies of *Tvv1* identified in the PN40024 sequence and in a BAC clone of Vidal blanc, we have deduced the structure of the canonical *Tvv1* element copy constituted by two perfect LTRs, 149 bp-long, flanking an ORF 4,146 bp-long that encode a putatively functional 1,382 amino acid-long polyprotein within a single translational reading frame. Two copies from the PN40024 genome sequence in addition to the *Tvv1*-VB were identified as canonical, the three differing in their UTL region sized 768–891 bp. The perfect identity between both LTRs indicates that these copies had recently inserted at their loci (<0.5 million years). No identical elements inserted at different loci were identified among the 22 full-length copies identified in the PN40024 genome. Thanks to the homozygosity level of this accession estimated at about 93% (The French–Italian Public Consortium for Grapevine Genome Characterization 2007), half of the *Tvv1* could be represented in this genome sequence, indicating that *Tvv1* family is constituted by a rather small number of elements, most of them being inactive.

Tvv1- Δ 3001 and *Tvv1*- Δ 3460, two deleted copies of *Tvv1* from cabernet sauvignon

Tvv1- Δ 3001 and *Tvv1*- Δ 3460 are two *Tvv1* elements that have undergone major rearrangements compared to *Tvv1*-VB used as reference. Both elements had suffered deletions leading to the early stopping of the polyprotein synthesis before the nucleic acid binding site of the *gag* domain. *Tvv1*- Δ 3001 shows a 3,001 bp-long deletion within the ORF in addition to a 33 bp-duplication within the UTL region. Another duplication, 32 bp-long, has previously been reported in the *Tvv1* fragment 3B2 at the end of the RNaseH domain (accession number AF478366; Pelsy and Merdinoglu 2002). Both duplications are in direct orientation, but while duplications of 3B2 are in tandem, ten nucleotides separate the duplication in the *Tvv1*- Δ 3001 UTL. *Tvv1*- Δ 3460 shows a larger deletion, 3,460 bp-long,

Table 4 Distribution of *Tvv1* full-length and deleted elements within 52 cultivated grapevine and six wild vines, *Vitis vinifera* ssp *silvestris*

<i>Accession full name</i>	<i>Tvv1</i>	<i>Tvv1</i> - Δ 3001	<i>Tvv1</i> - Δ 3460
	B-4000	B-1700	B-1500
aligoté B	+	–	+
aubin vert N	+	–	+
auxerrois B	+	–	+
bachet noir N	+	–	+
beaunois N	+	–	+
cabernet franc N	+	–	+
cabernet sauvignon N	+	+	+
carignan N	+	–	+
chardonnay B	+	–	+
chasselas blanc B	+	–	+
chenin B	+	–	+
cinsaut N	+	+	+
clairette B	+	–	+
colombard B	+	–	+
corbeau N	+	–	+
côt N	+	–	+
folle blanche B	+	–	+
franc noir de la Haute Saone N	+	–	+
gamay blanc Gloriod B	+	–	+
gamay noir N	+	–	+
gewurztraminer B	+	–	+
gouais B	+	–	+
grenache N	+	+	+
knipperlé B	+	–	+
marsanne B	+	–	+
mauzac B	+	–	+
melon B	+	–	+
merlot N	+	–	+
mourvèdre N	+	–	–
muscat cendré B	+	+	+
muscat d’Alexandrie B	+	–	+
muscat d’Alsace rouge N	+	+	+
muscat de Hambourg N	+	–	+
muscat Ottonel B	+	+	+
muscat à petits grains B	+	+	+
muscat reine des vignes B	+	–	+
muscat de Saumur B	+	–	+
persan N	+	–	+
peurion N	+	–	+
pinot noir N	+	–	+
riesling B	+	–	+
romorantin N	+	–	+
roublot N	+	–	+
roussanne B	+	–	–
sacy B	+	–	+
sauvignon blanc B	+	+	+

Table 4 continued

Accession full name	<i>Tvv1</i>	<i>Tvv1-Δ3001</i>	<i>Tvv1-Δ3460</i>
	B-4000	B-1700	B-1500
sémillon B	+	–	–
sylvaner B	+	+	+
syrah N	+	–	+
tannat N	+	–	+
ugni blanc B	+	–	+
vioigner B	+	–	+
<i>Vitis vinifera ssp silvestris</i> 50K	+	–	+
<i>Vitis vinifera ssp silvestris</i> 50L	+	–	+
<i>Vitis vinifera ssp silvestris</i> 53I	+	–	+
<i>Vitis vinifera ssp silvestris</i> 53J	+	–	+
<i>Vitis vinifera ssp silvestris</i> C25S2B	+	+	+
<i>Vitis vinifera ssp silvestris</i> C25S6	+	–	+

Cultivated grapevine types are white (B) or black (N)

encompassing the *Tvv1- Δ 3001* one. Moreover, *Tvv1- Δ 3460* UTL (1,038 bp) mainly differs from the *Tvv1-VB* UTL (768 bp) by insertions of five stretches as previously reported to explain size variation of the *Tvv1* UTL regions (Pelsy 2007).

The deleted copies of *Tvv1- Δ 3001* and *Tvv1- Δ 3460* are monogenic traits

Segregation of B-1700 and B-1500 shows that only one copy of *Tvv1- Δ 3001* and one of copy of *Tvv1- Δ 3460* are present in the muscat Ottonel and in the cabernet sauvignon genome, assuming that the observed segregation bias for B-1700 in the cabernet sauvignon self-progeny results from the low number of seedlings available. As they have not been spread around the grapevine genome, these two deleted copies cannot be considered as members of a subgroup of multicopy elements that lack protein-coding capacity such as TRIMs, *Morgane* or LARDs, whose mobility depends on trans-complementation by products of autonomous retrotransposons (Kalendar et al. 2004).

Hence, the difference observed between the sequenced A-1200, B-1700 A-750 and B-1500 fragments that were amplified from a single template must probably result from experimental errors (average mismatch 1/770 bp).

Formation of the deleted copies of *Tvv1- Δ 3001* and *Tvv1- Δ 3460* by illegitimate recombination

Retrotransposons often serve as substrates for recombination. Thanks to LTRs homology, recombination can occur

between two LTRs of integrated retrotransposons through double-strand break repair. Such events lead to deletions leaving solo LTRs (Shirasu et al. 2000). In *Tvv1- Δ 3001* and *Tvv1- Δ 3460*, the deletion breakpoint is marked by a stretch sized 13 and 11 bp, respectively, far shorter than common LTR sequences. The same short stretches were found almost identically in the canonical *Tvv1* flanking the deleted sequence. This feature provides an argument, albeit indirect, in favor of the deletions resulting from illegitimate homologous recombination events.

The deletion events leading to *Tvv1- Δ 3001* and *Tvv1- Δ 3460* could have occurred by recombination within one integrated *Tvv1* copy or between two independently inserted *Tvv1* copies. Research in *Arabidopsis* and tobacco (Gorbunova and Levy 1997; Kirik 2000) showed that strand rejoining after a break frequently occurs at short repeats and results in deletion from a few base pairs to several kilobases of DNA, the average deletion size depending on the species and being significantly shorter in tobacco than in *Arabidopsis* (920 vs. 1,341 bp; Kirik 2000). Such recombination events between identical stretches of two *Tvv1* copies can however be excluded because both deleted retrotransposons are flanked by perfect host repeats that mark specifically each element insertion. Recombination between identical stretches within a single *Tvv1* copy would have implied the presence of a full-length element prior to the recombination event. This second hypothesis can also be excluded because the insertion sites of *Tvv1- Δ 3001* and *Tvv1- Δ 3460* in the grapevine varieties studied either contain the deleted element or are empty.

The deletion events could also have occurred prior to integration. A source of genetic variation for viruses results in the exchange of genetic information facilitated by the presence of two RNA genomes in the same particle (Hu and Temin 1990). During retrovirus replication, the two RNA molecules packaged as a dimer can form a local heteroduplex at a region of self-complementarities, facilitating the generation of homologous recombinants. Recombination mainly occurs during original minus strand DNA synthesis and a misalignment of the two RNA molecules, thanks to repeated sequences, can result in a deleted copy (Temin 1993). In Brome mosaic virus (BMV) a nucleotide identity as short as 15 nucleotides can support efficient homologous recombination events (Nagy and Bujarski 1995). Retrotransposons such as Ty1 elements are known to recombine during retrotransposition (Boeke et al. 1986). In the same way as for virus, recombination can take place in the VLP of retrotransposons between the two packaged RNA molecules prior to the insertion of the newly synthesized DNA molecule. Such a mechanism could explain the formation of the *Tvv1* deleted copies. Supporting this idea, translation of

the sequences upstream and downstream of the deletion of *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640 leads to the original ORF without stop codon, which is expected if the two deleted elements were formed from autonomous copies of *Tvv1*. Marillonnet and Wessler (1998) have put forward this hypothesis to explain the structure of *wxStonor*, which suffered a deletion including the full *gag* region and a portion of the protease domain of a *Stonor* element, but leaving intact the remaining ORF. However, an extensive study of all *Stonor* members showed that most of their structural variation was mainly restricted to the region just downstream of the 5'LTR. Therefore, the authors proposed that the *wxStonor* deletion occurred during retrotransposition thanks to hybridization between extra nucleotides resulting from the addition of the tRNA in the PBS region and a homologous sequence downstream of the deletion (Marillonnet and Wessler 1998). Contrary to *wxStonor*, *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640 deletions span the ORF, leaving part of the *gag* region upstream and of the RnaseH domain downstream. In addition, the deletion breakpoints of *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640 are characterized by 13 or 11 bp-long stretches, which are duplicated in the canonical copy at both sides of the deleted regions (Table 3). All these results lead us to propose a model explaining the generation of *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640 that implies illegitimate recombination within one RNA molecule of a full-length *Tvv1* element or between two RNA molecules co-packaged in the VLP during reverse transcription, prior to the integration of the deleted copy into the host genome. Thanks to the two short duplicated stretches, the RNA molecule(s) could have misaligned, then led to template switching during minus strand DNA synthesis and finally to the synthesis of the deleted element. Based on the similarity of their UTL sequences, *Tvv1*- Δ 3001 would derive from a copy very similar to the putative active copy *Tvv1*-VB, while *Tvv1*- Δ 3640 displaying a different UTL sequence would have arisen from a different master copy of *Tvv1*.

Chronology of the *Tvv1*- Δ 3001 and *Tvv1*- Δ 3640 insertions

Flanking regions of *Tvv1*- Δ 3640 were shown to be identical for three varieties, cabernet sauvignon, pinot noir and muscat Ottonel. This result excludes that, at least in these varieties, different *Tvv1*- Δ 3640 copies were issued from independent recombination events and inserted at different sites. Hence, we assert that *Tvv1*- Δ 3640 is one fixed copy resulting from a unique recombination event. Its presence in most of the varieties tested indicates that it has most likely been dispersed through natural intermixing. However, in the three varieties, *Tvv1*- Δ 3640 LTRs differed in

the number of TA motifs. We propose that variation in the number of microsatellite repeats occurred subsequent to integration.

The dispersion by crosses of a single-locus deleted element is supported by pedigree reconstitution: *Tvv1*- Δ 3001 is present in ten varieties; some of them related as previously shown by microsatellite analysis. For example, cabernet sauvignon, a sibling from cabernet franc and sauvignon blanc (Bowers and Meredith 1997), shares *Tvv1*- Δ 3001 with sauvignon blanc, and muscat cendré has inherited *Tvv1*- Δ 3001 from muscat Ottonel (Bronner 2003).

Distribution of B-1500 and B-1700 is very different: while most of the varieties and wild vines (55/58) display *Tvv1*- Δ 3460, only ten varieties display *Tvv1*- Δ 3001. This result can be explained by the chronology of the insertions of *Tvv1*- Δ 3001 (\approx 1 million years) being more recent than *Tvv1*- Δ 3460 (\geq 2.4 millions years) that is older, more widely distributed and the sequence of which is less conserved. It is in agreement with the fact that we have not found any accessions that displayed *Tvv1*- Δ 3001, but not *Tvv1*- Δ 3460, and also with paleo-botanical data that record the first Vitaceae fossils from the Miocene (23.0–5.3 millions years; Galet 1988).

Acknowledgments The authors thank Dr. Eva Zyprian and Dr. Reinhardt Töpfer from the Federal Centre for Breeding Research on Cultivated Plant, Institute for Grapevine Breeding Geilweilerhof (Siebeldingen, Germany) for kindly providing the partial sequence of the BAC clone A1-16, Riccardo Velasco from the Istituto Agrario San Michele all'Adige (Italy) for the pinot sequence (VV78X150312.9) and Anne Françoise Adam-Blondon (INRA-URGV, France) for the BAC clone 3E11. We thank Lucie Bevilacqua-Schehrer and Cindy Ponama for their valuable technical assistance, Vincent Dumas and the experimental domain staff for grapevine accessions, Malek Alioua (ULP-Strasbourg, France) for sequencing, and Pere Mestre for careful revision of the English. Cedric Moisy is a doctoral fellow supported by the Conseil Régional d'Alsace. All experiments described in this study comply with the current laws of France.

References

- Anonymous (1995) Catalogue des variétés et clones de vigne cultivés en France, ENTAV, Ed, Le Grau du Roi. France, p 357
- Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. *Ann Rev Microbiol* 43:403–434
- Boeke JD, Styles C, Fink R (1986) *Saccharomyces cerevisiae* *SPT3* gene is required for transposition and transpositional recombination of chromosomal Ty elements. *Mol Cell Biol* 6:3575–3581
- Bowers J, Meredith C (1997) The parentage of a classic wine grape, Cabernet Sauvignon. *Nature Genet* 16:84–87
- Bronner A (2003) Muscats et variétés muscatées. *Inventaire et synonymie universels, des origines à nos jours*, Avenir-Oenologie, INRA Editions/Oenoplurimedia, p 222
- Casacuberta JM, Grandbastien M-A (1993) Characterization of LTR sequences involved in the protoplast specific expression of the tobacco *Tnt1* retrotransposon. *Nucl Acids Res* 21:2087–2093

- Casacuberta JM, Vernhettes S, Grandbastien M-A (1995) Sequence variability within the tobacco retrotransposon *Tnt1* population. *EMBO J* 14:2670–2678
- Feng Y-X, Moore SP, Garfinkel DJ, Rein A (2000) The genomic RNA in Ty1 virus-like particles is dimeric. *J Virol* 74:10819–10821
- Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nature Rev Genet* 3:329–341
- Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucl Acids Res* 20:3639–3644
- Galet P (1988) Cépages et vignobles de France. Les vignes américaines, Edition, p 553
- Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proc Natl Acad Sci USA* 93:10274–10279
- Gorunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucl Acids Res* 25:4650–4657
- Hu W, Temin HM (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci USA* 87:1556–1560
- Jordan IK, McDonald JF (1998) Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. *J Mol Evol* 47:14–20
- Jordan IK, McDonald JF (1999) Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. *Genetics* 151:1341–1351
- Kalendar R, Vicent CM, Peleg O, Amamthawat-Jonsson K, Bolshoy A, Schulman A (2004) Large retrotransposon derivatives: abundant, conserved by non autonomous retroelements of barley and related genomes. *Genetics* 166:1437–1450
- Kirik A, Salomon S, Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO J* 19:5562–5566
- Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. *Science* 304:982
- Lamoureux D, Bernole A, Clainche IL, Tual S, Thareau V, Paillard S, Legeai F, Dossat C, Wincker P, Oswald M, Merdinoglu D, Vignault C, Delrot S, Caboche M, Chalhou B, Adam-Blondon A-F (2006) Anchoring of a large set of markers onto a BAC library for the development of a draft physical map of the grapevine genome. *Theor Appl Genet* 113:344–356
- Marillonnet S, Wessler SR (1998) Extreme structural heterogeneity among the members of a maize retrotransposon family. *Genetics* 150:1245–1256
- Nagy PD, Bujarski JJ (1995) Efficient system of homologous RNA recombination in brome mosaic virus: sequence and structure requirements and accuracy of crossovers. *J Virol* 69:131–140
- Nymark-McMahon MH, Beliakova-Bethell NS, Darlix J-L, Le Grice SFJ, Sandmeyer SB (2002) Ty3 integrase is required for initiation of reverse transcription. *J Virol* 76:2804–2816
- Pelsy F (2007) Untranslated leader region polymorphism of *Tvv1*, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*. *Theor Appl Genet* 116:15–27
- Pelsy F, Merdinoglu D (2002) Complete sequence of *Tvv1*, a family of Ty1 *copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking. *Theor Appl Genet* 105:614–621
- Preston BD (1996) Error-prone retrotransposition: Rime of the ancient mutators. *Proc Natl Acad Sci USA* 93:7427–7431
- Sabot F, Sourdille P, Chantret N, Bernard M (2006) *Morgane*, a new LTR retrotransposon group, and its subfamilies in wheats. *Genetica* 128:439–447
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20:43–45
- Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res* 10:908–915
- Suoniemi A, Tanskanen J, Schulman AH (1998) Gypsy-like retrotransposons are widespread in the plant kingdom. *Plant J* 13:699–705
- Temin H (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci USA* 90:6900–6903
- The French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463–467
- Verriès C, Bès C, This P, Tesnière C (2000) Cloning and characterization of *Vine-1*, a LTR- retrotransposon-like element in *Vitis vinifera* L., and other *Vitis* species. *Genome* 43:366–376
- Vicent CM, Kalendar R, Schulman AH (2005) Variability, recombination, and mosaic evolution of the barley *BARE-1* retrotransposon. *J Mol Evol* 61:275–291
- Vicent CM, Suoniemi A, Anamthamat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH (1999) Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11:1769–1784
- Voytas DF, Cumming MP, Konieczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* 89:7124–7128

Chapitre II : Caractérisation d'un rétrotransposon du génome de *S. demissum* présentant des similarités avec *Tvv1*

Pour évaluer la distribution de la famille *Tvv1* chez les plantes, nous avons recherché par Blast des séquences similaires à sa région *gag* dans les bases de données disponibles sur le site NCBI (recherche réalisée en 2006). Cette recherche nous a permis de mettre en évidence la séquence d'un clone BAC de *Solanum demissum*, une variété sauvage de pomme de terre, qui présente des motifs très similaires à la séquence requête. L'analyse de cette séquence et sa comparaison à la séquence de *Tvv1* ont permis la caractérisation d'un nouveau rétrotransposon du génome de *S. demissum*, que nous avons baptisé *Tvv1*-SD en raison de sa proximité à la famille *Tvv1* de la vigne.

II.1.1 - Résultats

a) Description de l'élément *Tvv1*-SD

La recherche par Blast a été réalisée en utilisant comme requête une séquence de 215 acides aminés correspondant à la région *gag* de la polyprotéine de la copie *Tvv1*-VB (Moisy *et al.*, 2008). Le rétrotransposon *Tvv1*-SD a été détecté entre les nucléotides n°13231 et 18051 du clone BAC n° AC149291 de *S. demissum* (NCBI). Sa séquence de 4821 pb est bordée de chaque côté par la duplication de la séquence CTCGA de l'hôte. La région interne de 4461 pb est encadrée par deux LTRs de 180 pb chacune qui se différencient par 3 nucléotides (98,3% d'identité) (Figure 51). La date d'insertion de l'élément dans le génome de *S. demissum* a été estimée à environ 1,3 millions d'années, en utilisant la méthode d'évaluation proposée par Vitte et collaborateurs (2007).

La séquence de *Tvv1*-SD contient un cadre de lecture ouvert qui débute par un codon ATG situé 214 pb après la LTR 5'. En considérant la présence d'un saut dans le cadre de lecture (frameshift +1), cet ORF code potentiellement pour une polyprotéine de 1413 acides aminés (Figure 51). Cette polyprotéine présente tous les domaines attendus chez les rétrotransposons

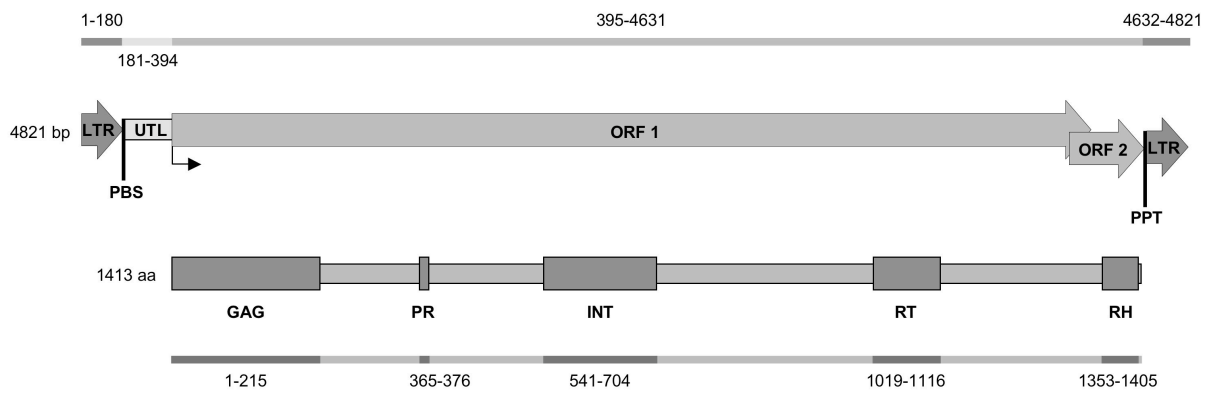


Figure 51 : Structure détaillée de la séquence l'élément *Tvv1*-SD de *S. demissum*.

de type Ty1-*copia*-like, c'est-à-dire les domaines *gag*, protéase, intégrase, transcriptase inverse et RNaseH, disposés dans cet ordre le long de la séquence. La séquence TGAGGGGGAG, correspondant au motif PPT (Poly-Purine Tract) est présente en amont de la LTR 3'. La région interne située entre la LTR 5' et le début de la région codante représente une séquence UTL de 214 pb. Elle débute par la séquence TGGTATCAGAGCC, qui correspond au site PBS (Primer-Binding Site) complémentaire de l'ARNt méthionine.

b) Relations phylogénétiques avec les autres rétrotransposons de plantes

Afin d'évaluer l'appartenance de *Tvv1*-SD à la famille *Tvv1* de vigne, plusieurs séquences nucléiques de rétrotransposons de plantes ont été utilisées pour construire un arbre phylogénétique à l'aide du logiciel MEGA 3.1 (Kumar *et al.*, 2004), en utilisant la méthode de Neighbor-Joining et le modèle de Jukes et Cantor. Les rétrotransposons qui ont été utilisés sont les suivants : *AtRE1* (n° accession : AB021265), *BARE-1* (Z17327), *Opie-2* (AF090446), *PDR1* (X66399), *PREM-2* (U41000), *Retrolyc1-1* (AF228701), *Rtsp-1* (AB162659), *Tal* (X13291), *TLC1-1* (AF279585), *Tnt1* (X13777), *Tst1* (X52387), *Tto-1* (D83003), *RIRE-1* (D85597), *Wis-2-1A* (AF339051), ainsi que les rétrotransposons de vigne décrits dans la Partie II de ce manuscrit.

D'autre part, les motifs I à VII de la région transcriptase inverse des polyprotéines codées par ces éléments ont été isolés puis comparés entre eux. Lorsque les séquences protéiques n'étaient pas disponibles, la région interne des rétrotransposons a été traduite selon les trois cadres de lecture possibles, et la séquence protéique de la région transcriptase inverse a été reconstruite manuellement.

L'alignement de la séquence nucléique complète de l'élément *Tvv1*-SD avec celles de rétrotransposons de plantes a permis la construction d'un arbre phylogénétique qui tient compte de la variabilité observée sur la totalité de la séquence des rétrotransposons (Figure 52). La position de l'élément *Tvv1*-SD dans cet arbre permet de le distinguer de tous les rétrotransposons déjà décrits chez les plantes, et notamment de ceux décrits chez les *Solanaceae* : les éléments *Retrolyc1-1* (tomate), *TLC1-1* (*L. chilense*, tomate sauvage), *Tnt1* et *Tto-1* (tabac) et *Tst1* (*S. tuberosum*, la pomme de terre cultivée). La proximité de *Tvv1*-SD avec la famille *Tvv1* contraste avec l'éloignement de leurs espèces hôtes respectives, la pomme de terre et la vigne, illustré par la Figure 53 (Brunner *et al.*, 2004).

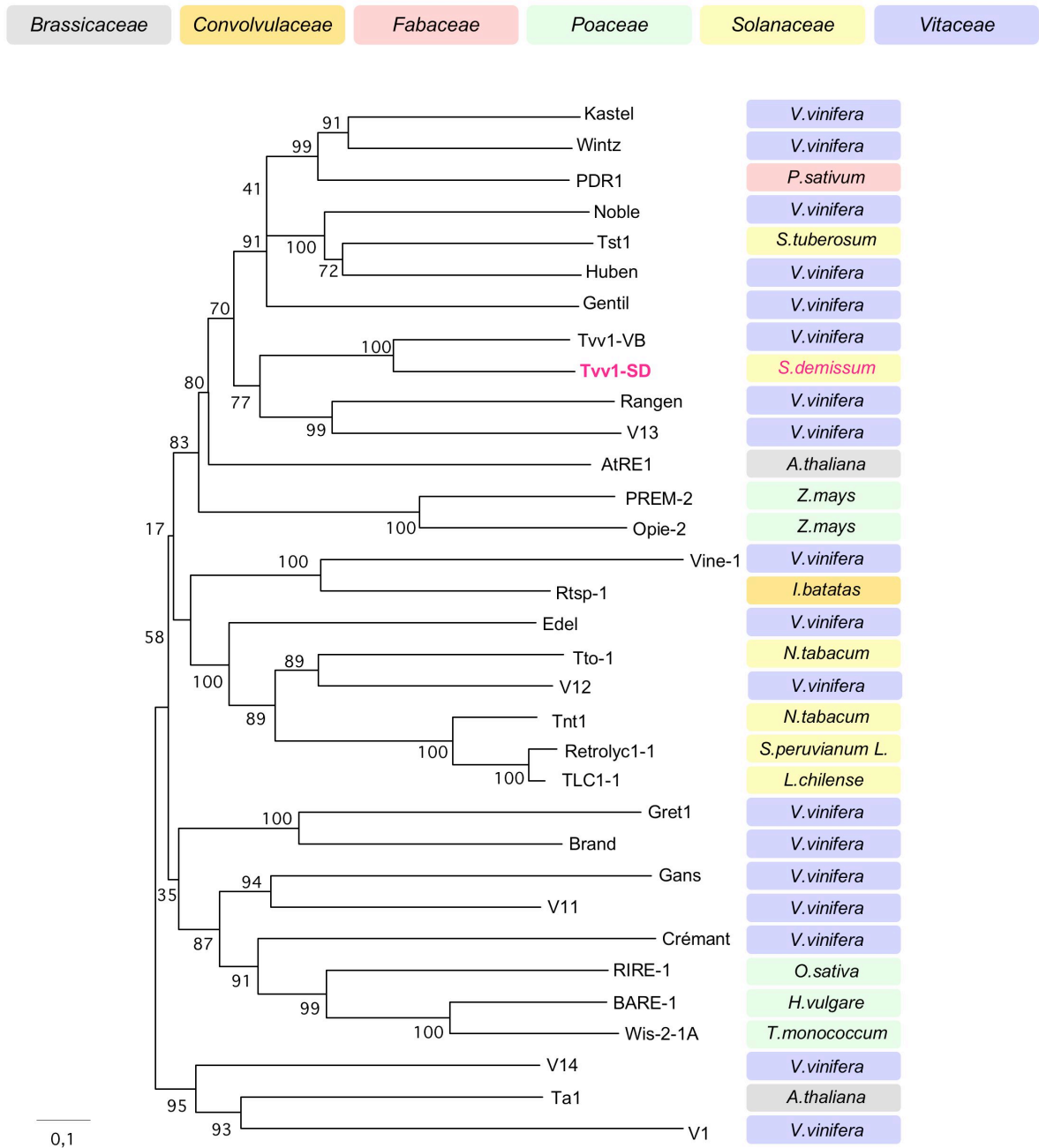


Figure 52 : Arbre phylogénétique de rétrotransposons connus chez les plantes (Neighbor-Joining, 1000 répétitions bootstrap, distances estimées avec le modèle Jukes et Cantor).

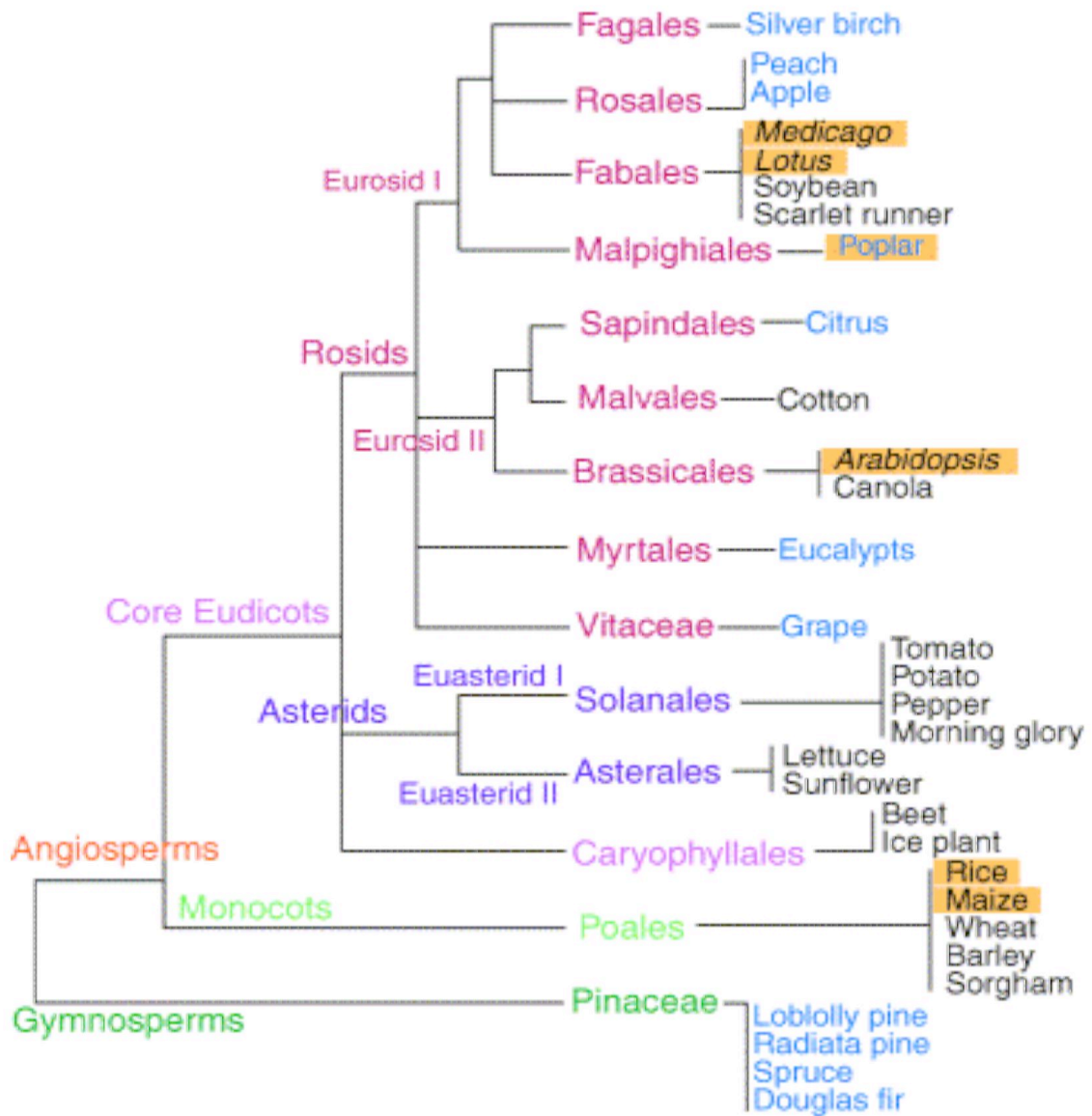


Figure 53 : Classification des espèces végétales (d'après Brunner, 2004).

	1		35
LTR 3' Tvv1-VB	T G T T A - - - - -	G C T G T A T A T A T - C T G T - - - - -	
LTR 5' Tvv1-VB	T G T T A - - - - -	G C T G T A T A T A T - C T G T - - - - -	
LTR 5' Tvv1-SD	T G T G A C A A A A T	G T T G T A A A T A T T C T A T T A C C T A A T	
LTR 3' Tvv1-SD	T G T G A C A G A A T	G T T G T A A A T A T T C T A T T A C C T A A T	
	36		70
LTR 3' Tvv1-VB	- - - - - A C A - T A C C A T A - - - - -	A T T T G G T T G T T	
LTR 5' Tvv1-VB	- - - - - A C A - T A C C A T A - - - - -	A T T T G G T T G T T	
LTR 5' Tvv1-SD	A T A G T C A G C A G T A T C A T A G G A A A A	A T C T G T A A G A A	
LTR 3' Tvv1-SD	A T A G T C A A C A G T A T C A T A G G A A A A	A T C T G T A A G A A	
	71	TATA Box	105
LTR 3' Tvv1-VB	T C - - - - - - - - -	C T T A G G G A T A A T A C C T T C C - - T	
LTR 5' Tvv1-VB	T C - - - - - - - - -	C T T A G G G A T A A T A C C T T C C - - T	
LTR 5' Tvv1-SD	T C A T A A G A G A G A T	C T G T A G G A T A T T T C T T A C C T T	
LTR 3' Tvv1-SD	T C A T A T G A G A G A T	C T G T A G G A T A T T T C T T A C C T T	
	106	Limite U3 / R	140
LTR 3' Tvv1-VB	A A T T C A G G A C T C T C A A T T G T A T A T A T A A A C A A G T A		
LTR 5' Tvv1-VB	A A T T C A G G A C T C T C A A T T G T A T A T A T A A A C A A G T A		
LTR 5' Tvv1-SD	T A G T A T A G C T T C T T G C T A G T G T A T A T A T A C A T G T A		
LTR 3' Tvv1-SD	T A G T A T A G C T T C T T G C T A G T G T A T A T A T A C A T G T A		
	141	Site Polyadénylation	175
LTR 3' Tvv1-VB	T T A T T C A T C T A A T A A A G A C A A G G A A T T G A G A A T T A		
LTR 5' Tvv1-VB	T T A T T C A T C T A A T A A A G A C A A G G A A T T G A G A A T T A		
LTR 5' Tvv1-SD	A T G T T T T A T A T A T G A A A T A A G A A G T T C T G - - - -		
LTR 3' Tvv1-SD	A T G T T T T A T A T A T G A A A T A A G A A G T T C T G - - - -		
	176	191	
LTR 3' Tvv1-VB	C C T T G A T T C G G T T A C A		
LTR 5' Tvv1-VB	C C T T G A T T C G G T T A C A		
LTR 5' Tvv1-SD	- - - - - T T C A A T T A C A		
LTR 3' Tvv1-SD	- - - - - T T C A A T T A C A		

Figure 54 : Comparaison de la séquence des LTRs des éléments *Tvv1*-SD et *Tvv1*-VB.

c) Comparaison de l'élément de *S. demissum* à ceux de la famille *Tvv1* de vigne

La taille des régions LTRs de l'élément *Tvv1*-SD de *S. demissum* (180 pb) est du même ordre de grandeur que la taille des LTRs des éléments de la famille *Tvv1* de vigne, comprises entre 149 et 198 pb (Pelsy and Merdinoglu, 2002; Moisy *et al.*, 2008). De plus, les LTRs du nouvel élément présentent 46% d'identité avec celles de *Tvv1*-VB, utilisé comme élément de référence. L'alignement des séquences LTR montre que les premiers et les derniers nucléotides (respectivement TGT et TTACA) sont conservés, ainsi que d'autres blocs de 5 à 66 pb dont les pourcentages d'identité sont compris entre 53 et 80% (Figure 54). La limite des régions U3 et R, décrite pour l'élément *Tvv1* de la vigne (Pelsy and Merdinoglu, 2002), est présente chez *Tvv1*-SD. Le pourcentage de bases conservées entre les LTRs des deux éléments dépend de la région considérée. En considérant le nombre et la taille des *gap* (respectivement 7 et 1 *gap* de 1 à 15 pb), ainsi que sur le pourcentage de nucléotides conservés chez les deux éléments, la région U3 est sensiblement plus divergente que la région R. Les motifs d'initiation de la transcription (TATA box) et de poly-adénylation, précédemment décrits pour la famille *Tvv1* de la vigne (Pelsy and Merdinoglu, 2002), ne sont pas complètement conservés dans les LTRs de l'élément *Tvv1*-SD (Figure 54).

L'alignement des deux polyprotéines potentiellement codées par *Tvv1*-SD (1413aa) et *Tvv1*-VB (1382aa) montre que leurs séquences présentent un taux d'identité et de similarité de 69% et 79%, mais que ces taux varient fortement selon le domaine considéré (Figure 55). En effet, la région *gag* de la polyprotéine (215aa) est identique à 67,4% (87% de similarité) entre les deux éléments et les régions intégrase (52aa) et RNaseH (64aa) sont identiques à 82,7% et 85,9% respectivement. La séquence comprise entre les motifs I et VII de la région transcriptase inverse présente un taux d'identité et de similarité de 81,5% et 90,3% entre les éléments *Tvv1*-SD et *Tvv1*-VB. Comme nous l'avons vu précédemment, un taux d'identité supérieur à 79% sur cette séquence permet de regrouper les éléments dans une même famille de rétrotransposons du génome de la vigne (voir article en cours de préparation, page 63). Aussi, la conservation de ce domaine entre les éléments *Tvv1*-SD et *Tvv1*-VB confirme leur appartenance à la même famille.

Plus particulièrement, la comparaison des séquences nucléiques et protéiques des deux éléments a permis de délimiter, en amont du domaine RNaseH, une délétion de 35 pb sur la séquence de l'élément *Tvv1*-SD, qui provoque le décalage du cadre de lecture (frameshift +1)

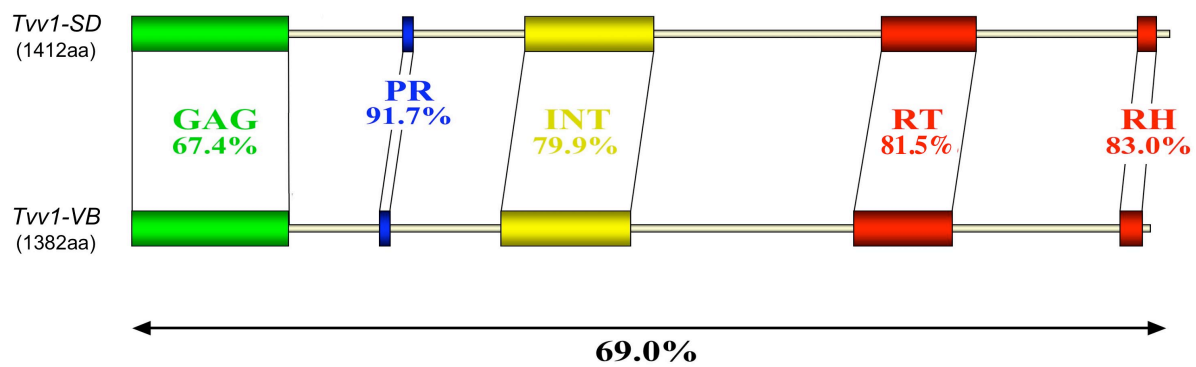


Figure 55 : Taux d'identité observés entre les polyprotéines potentiellement codées par les éléments *Tvv1-SD* de *S. demissum* et par *Tvv1-VB* de *V. vinifera*.

et empêche probablement la synthèse d'une polyprotéine complète chez *S. demissum* (Figure 51).

II.1.2 - Conclusions

Le rétrotransposon présent dans le clone BAC de *Solanum demissum* possède tous les domaines attendus chez un élément de type Ty1-*copia*-like. La présence de séquences de l'hôte dupliquées de part et d'autre de cet élément et l'estimation de la date d'intégration de la copie à environ 1,3 Ma prouvent que l'élément *Tvv1*-SD s'est inséré relativement récemment dans le génome de *S. demissum*.

En considérant le décalage du cadre de lecture probablement provoqué par la délétion de 35 pb, sa région interne code potentiellement pour une polyprotéine complète, portant tous les domaines nécessaires à l'autonomie de la copie. Cependant, la présence de ce décalage et l'absence des motifs d'initiation et de poly-adénylation, qui sont indispensables à la transcription d'un ARN complet, laissent supposer que cette copie n'est probablement pas autonome. Néanmoins, des signaux régulateurs différents de ceux de l'élément *Tvv1* de vigne et spécifiques de l'élément de la pomme de terre sont peut-être présents dans les LTRs.

La position de l'élément *Tvv1*-SD dans l'arbre phylogénétique des rétrotransposons de plante permet de le distinguer très nettement des éléments précédemment décrits chez les *Solanaceae* et le place à proximité de la famille *Tvv1*. La comparaison de sa séquence protéique complète avec celle de *Tvv1*-VB, la copie de référence de la famille *Tvv1*, montre que leurs différents domaines fonctionnels sont très similaires, notamment pour la région *gag*, qui est pourtant considérée comme la région la moins conservée entre les familles de rétrotransposons. Le domaine transcriptase inverse, qui est couramment utilisé pour évaluer l'appartenance des rétrotransposons à une même famille, est identique à plus de 80% entre *Tvv1*-SD et *Tvv1*-VB, ce qui prouve qu'ils appartiennent à la même famille. Les LTRs, décrites comme les régions qui évoluent les plus rapidement chez les rétrotransposons (Lankenau *et al.*, 1990; Kalmykova *et al.*, 2004), sont relativement proches entre les deux éléments. Ainsi, il est possible que ces deux rétrotransposons aient une origine commune malgré la distance importante qui sépare leurs espèces hôtes, la pomme de terre et la vigne.

Pour expliquer la présence de ces deux éléments chez des espèces végétales aussi éloignées, deux hypothèses peuvent être envisagées.

Tout d'abord, les deux éléments peuvent être issus d'un même et unique élément « ancestral ». Cet élément, présent dans le génome d'un ancêtre commun aux deux espèces-hôte, se serait maintenu dans les deux génomes après la différenciation des deux espèces. Dans ce cas, la proximité actuelle des deux rétrotransposons pourrait s'expliquer par le fait que leur structure et leur séquence aient été conservées au cours de l'évolution, soit en raison de leurs capacités fonctionnelles qui leur ont permis « d'échapper » au contrôle de l'hôte, soit parce que ces deux éléments n'ont pas eu d'effet néfaste sur la fitness du génome hôte et n'ont donc pas été contre-sélectionnés, soit parce que leur présence joue un rôle particulier dans le fonctionnement du génome. Leurs séquences auraient ainsi relativement peu divergé l'une de l'autre.

Dans une seconde hypothèse, la présence des deux rétrotransposons chez ces espèces végétales pourrait résulter d'un événement de transfert horizontal de la vigne vers la pomme de terre, ou inversement. Dans ce cas, l'un des deux éléments aurait été « transmis » au nouveau génome hôte : le croisement de la vigne et de la pomme de terre étant très peu probable, il est possible qu'un des éléments ait été « transmis » par l'intermédiaire d'un pathogène et une fois insérée dans le nouveau génome, le rétrotransposon pourrait alors s'être multiplié. Plusieurs preuves de transferts horizontaux d'éléments transposables ont été apportées, notamment pour le rétrotransposon *RIRE1* chez le riz (Roulin *et al.*, 2007) ou le transposon *Mu* du maïs (Diao *et al.*, 2006), mais aussi chez la drosophile (Daniels *et al.*, 1990; Sanchez-Garcia *et al.*, 2005; Kotnova *et al.*, 2007). Cependant, la présence de *Tvv1*-SD n'a été détectée que dans un seul clone BAC de *S. demissum*, et d'autres Blast n'ont pas permis de mettre en évidence sa présence chez la pomme de terre cultivée *S. tuberosum*, ce qui pourrait vouloir dire qu'il n'est présent qu'en une seule copie dans le génome de *S. demissum*. Des amplifications PCR réalisées à partir d'ADN de *S. tuberosum*, la pomme de terre cultivée, en utilisant des amorces dessinées dans la séquence de *Tvv1*-SD nous ont permis d'obtenir des signaux aux tailles attendues (données non présentées). Si ces fragments n'ont pour le moment pas été séquencés, ils pourraient néanmoins confirmer la présence de *Tvv1*-SD dans d'autres espèces de pomme de terre, et donner une indication sur le nombre de copies.

Aucun élément proche de *Tvv1*-SD n'a pour le moment été détecté dans les espèces différentes de la pomme de terre et de la vigne. L'élément *Tvv1*-SD n'est présent ni dans le

génomique du tabac (Maunoury, INRA de Versailles, communication personnelle), ni dans les autres dicotylédones qui sont pourtant plus proches de la pomme de terre que ne l'est la vigne. Aussi, si l'hypothèse d'un transfert horizontal est possible, il serait alors probable que l'élément de vigne ait été transmis à la pomme de terre, et non l'inverse.

Partie V : Matériels et méthodes

Matériels et méthodes

Obtention de boutures

Ce protocole, adapté de la thèse de Julliard, intitulée « Etude physiologique de la rhizogenèse et des conditions de réussite du greffage de la vigne » (1973, Paris VI), est régulièrement utilisé à l'INRA de Colmar pour obtenir des boutures en bois.

Précautions préalables

Protéger le bois de tout dessèchement au moment du débitage et de la préparation des boutures, à l'aide d'une bâche plastique. Travailler lot par lot pour éviter tout risque de mélange de génotypes, utiliser des cuvettes pour manipuler chaque lot de bouture durant leur préparation. Avant de couper les boutures, nettoyer le bois de ses vrilles et entre cœurs.

Débiter les boutures

Une bouture talonnée est constituée d'un bois de un an aoûté, d'une longueur moyenne optimale de 15 cm (minimum de 10 cm) devant intégrer au moins un entre-nœud ou mérithalle (elle comporte ainsi au minimum deux nœuds). Soigner les sections (supérieure et inférieure) de la bouture (Figure 56). En aucun cas la section coupée ne doit être au niveau du bourgeon, laisser 5 à 10 mm de bois au-dessus du nœud. En bas, la section doit être positionnée à 4-5 mm en dessous du nœud. Le bourgeon du nœud supérieur est conservé, il assurera le développement végétatif de la bouture. Le(s) bourgeon(s) inférieur(s) est (sont) supprimé(s) au sécateur, c'est à leur niveau qu'aura lieu la rhizogenèse. Durant leur manipulation, il est préférable de regrouper les boutures de chaque génotype dans une cuvette avec l'étiquette du prélèvement. Les boutures sont regroupées par lots, à l'aide d'élastiques, avec leurs étiquettes respectives toujours fixées à l'une d'entre elles.

Paraffinage

L'extrémité supérieure de la bouture, c'est-à-dire l'œil laissé pour le développement, est paraffinée pour la protéger du dessèchement. Pour cela, la bouture est plongée dans un bain de cire (PP140F), pour que l'œil soit immergé entièrement et jusqu'à 5 cm en dessous. Faire fondre la cire dans un bac à paraffiner. Le point de fusion de la cire se situe entre 71 et 73°C. Dans la pratique, la température d'utilisation doit se situer entre 78 et 85°C. Le temps de trempage ne doit pas dépasser 1 seconde. Il est impératif de contrôler la température pendant

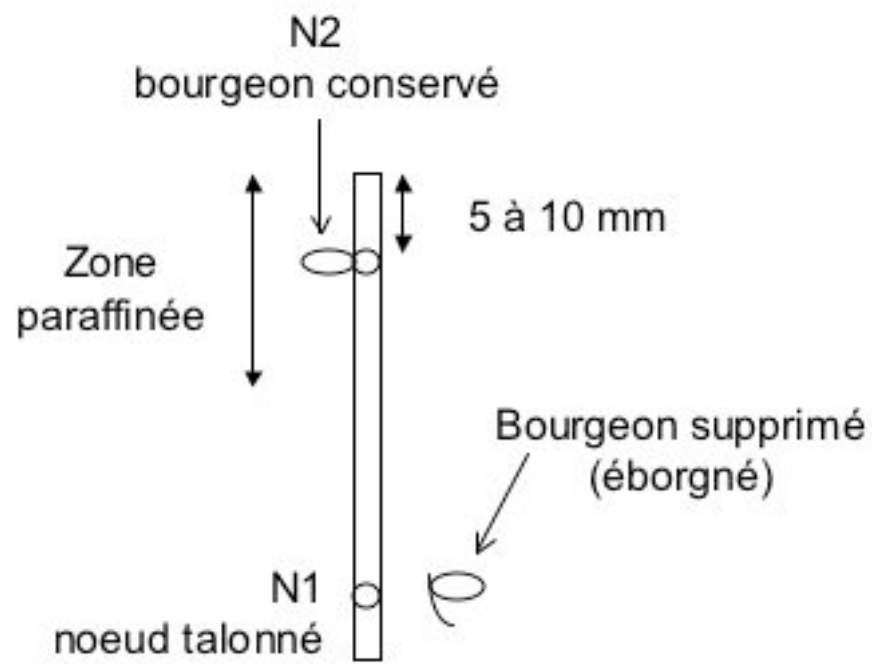


Figure 56 : Schéma de la découpe des sarments de vigne en boutures

toute la durée de l'utilisation. La pellicule de cire sur la bouture doit être nette, ni trop épaisse, ni trop fine. La bouture est trempée dans de l'eau froide pour solidifier immédiatement la cire.

Forçage

On appelle « forçage » le fait de forcer le développement des racines et du bourgeon des boutures en serre. Pour garantir une bonne rhizogenèse, la partie talonnée de la bouture est plongée dans une poudre d'hormone Acide 3-Indol Butyrique de type Chryzoteck beige, puis elle est agitée pour éliminer l'excédent de poudre car la quantité d'hormone apportée doit être très faible. La bouture est ensuite plantée dans un pot contenant un mélange 2/3 de sable et 1/3 de perlite (rapport volumique) en ne laissant dépasser que la partie couverte de cire. Les pots sont drainés par un horizon de 1 à 2 cm de gravier disposé au fond. Les plantes sont ensuite régulièrement arrosées d'eau. Les premiers signes de développement apparaissent quelques jours après, par exemple le gonflement des bourgeons et le débourrement.

Résultat

Les résultats dépendent essentiellement de la qualité de l'aoûtement et des réserves du bois. En principe, lorsque le bois vient de l'extérieur, le taux de réussite atteint 90 %. Par contre, le bois de qualité inférieure ne permet pas d'obtenir de si bons résultats.

Extraction des ARN totaux

Les extractions d'ARN totaux ont été réalisées sur des échantillons représentant au total 100 mg de matière végétale, en utilisant le kit RNeasy Plant Mini Kit (QUIAGEN). Plusieurs étapes du protocole ont été modifiées par rapport aux recommandations du fournisseur afin d'optimiser le rendement et la qualité des extractions.

Précautions préalables

Avant toute manipulation, les ustensiles nécessaires à l'extraction sont préalablement nettoyés à la soude (NaOH 1N) et au SDS-EDTA (5g/L, 0,4g/L), puis rincés avec de l'eau traitée au DEPC, emballés dans du papier aluminium et finalement autoclavés.

Broyage

À l'aide d'un mortier et d'un pilon, 100 mg de matière végétale sont broyés dans de l'azote liquide jusqu'à l'obtention d'une poudre très fine qui est transférée dans un tube 1,5 ml

contenant le tampon d'extraction (1 mL RLC, 25 mg PVP 40, 10 μ L de Beta-mercaptoéthanol).

Purification des Acides ribonucléiques

Le tube est vortexé 10 secondes puis le mélange est déposé en deux temps sur la colonne violette (fournie avec le kit) et centrifugé 2 min à 14000 rpm à température ambiante. Cette étape permet d'éliminer les débris cellulaires et les protéines végétales. Le surnageant est conservé à température ambiante, dans un tube 2 mL. Pour précipiter les acides nucléiques, 500 μ L d'éthanol 100% à -20°C sont ajoutés. Le mélange est pipeté et refoulé doucement, puis déposé sur la colonne rose (fournie avec le kit) et centrifugé 15 sec à 10000 rpm à température ambiante. La membrane de la colonne rose permet la fixation des ARN, l'éluat est donc éliminé. Pour laver la membrane, 350 μ L de tampon RW1 sont déposés sur la colonne, puis centrifugés pendant 15 sec à 10000 rpm à température ambiante.

Pour dégrader l'ADN résiduel, un mix contenant 10 μ L de DNaseI et 70 μ L de tampon RDD est déposé délicatement sur la membrane et laissé incubé 15 min à température ambiante. La membrane est lavée une seconde fois avec 350 μ L de tampon RW1. La colonne est alors posée sur un nouveau tube collecteur. Deux nouveaux lavages de la membrane sont réalisés avec 500 μ L de tampon RPE et une centrifugation de 15 sec et de 2 min respectivement, à 10000 rpm, à température ambiante. L'éluat est éliminé à chaque fois. La colonne est ensuite placée sur un tube de 1,5 mL et centrifugée 1 min pour éliminer toute trace d'éthanol résiduel susceptible de perturber l'élution des acides nucléiques. La colonne est posée sur un nouveau tube de 1,5 mL et 40 μ L d' H_2O RNase free sont déposés sur la membrane puis laissés incubés pendant 5 min à température ambiante. Une dernière centrifugation, pendant 1 min à 10000 rpm à 4°C , permet d'éluer les ARN totaux qui sont ensuite conservés à -20°C ou -80°C , selon leur utilisation.

Northern blot

La taille et l'accumulation des ARNm a été déterminée selon la méthode décrite par Lamotte et collaborateurs (Lamotte *et al.*, 2004). Dix microgrammes d'ARN totaux ont été dénaturés, séparés sur gel dénaturant puis transférés par capillarité sur une membrane de nylon Hybond-XL (Amersham-Pharmacia Biotech) selon le protocole décrit par Sambrook et collaborateurs (Sambrook *et al.*, 1989). La membrane a ensuite été préhybridée 3 h dans un tampon

(Na₂HPO₄ 500 mM pH 7,2 ; EDTA 1 mM ; BSA 1% ; SDS 7%) à 65°C, puis incubée 16 h à 65°C dans ce même tampon frais auquel a été ajoutée la sonde (produit PCR) radiomarquée au [α -³²P]CTP selon la méthode d'amorçage aléatoire (Prime-a-gene, Promega). La membrane est ensuite lavée 10 min sous agitation dans un tampon SSC 2X (SSC 1X : citrate de sodium 15 mM ; NaCl 220 mM ; pH 7,4) à température ambiante, puis dans un tampon SSC 2X, SDS 0,5% à 60°C pendant 1 h et finalement 10 min dans un tampon SSC 0,1X à température ambiante, avant autoradiographie.

Conclusion générale

Conclusion générale

La vigne cultivée est une plante pérenne propagée par multiplication végétative sur des pas de temps très longs, il en résulte que les variétés de vigne sont composées de clones morphologiquement très proches mais qui peuvent présenter des différences génétiques. En l'absence de reproduction sexuée, l'accumulation de mutations somatiques au cours des cycles de multiplication végétative représente l'unique source de variation génétique entre clones, et les éléments mobiles présents dans les génomes pourraient générer une part de cette variabilité.

Au début de ma thèse, les informations disponibles sur la diversité des éléments transposables du génome de la vigne étaient rares puisque seuls trois rétrotransposons à LTR étaient connus, les deux copies *Gret1* et *Vine-1*, ainsi que la famille *Tvv1*. Par ailleurs, aucune donnée n'était disponible sur leur niveau d'expression.

Afin d'apporter des éléments de compréhension du rôle des éléments transposables dans la genèse de la diversité intravariétale, mon projet de thèse a eu pour objectif d'identifier de nouvelles familles d'éléments transposables, puis d'en étudier la structure et la transcription.

Analyse structurale

Par une stratégie combinant l'analyse de fragments de séquences RNaseH-LTR isolés par la méthode de Pearce et une recherche par similarité de séquences dans le génome de PN40024, nous avons pu caractériser dix nouvelles familles de rétrotransposons à LTR de type Ty1-*copia*-like et identifier pour chacune d'elles des copies canoniques. Dans certaines familles, nous avons identifié des copies complètes, théoriquement capables de synthétiser les protéines nécessaires à leur cycle de rétrotransposition, et donc potentiellement autonomes. L'amplification de ces familles a certainement contribué à l'évolution de la taille du génome de la vigne et il semble qu'une partie de la quantité d'ADN apportée au génome par leur intégration ait été éliminée, probablement par des mécanismes de délétions ou de recombinaisons inter-éléments, comme cela a été décrit chez d'autres plantes (Sabot and Schulman, 2007; Vitte *et al.*, 2007b).

En prenant en compte les dix nouvelles familles que nous avons caractérisées, les éléments *Gret1*, *Tvv1*, *Vine-1* et cinq rétroéléments annotés par le Laboratoire Plant Genomics de l'Université d'Helsinki, ces familles représentent entre 4 et 15% des rétrotransposons de la vigne selon la méthode d'annotation considérée et occupent actuellement environ 2% du génome.

L'identification de copies *Tvv1* complètes nous a permis de valider la stratégie de marche sur le chromosome utilisée précédemment pour caractériser cette famille. Nous avons également pu déterminer que deux copies *Tvv1* délétées s'étaient probablement formées par un mécanisme de recombinaison survenu au cours de l'étape de transcription inverse.

La recherche de séquences similaires à *Tvv1* dans les bases de données nous a permis de caractériser un rétrotransposon dans le génome de la pomme de terre sauvage appartenant vraisemblablement à la famille *Tvv1* de la vigne. Compte tenu de la distance phylogénétique entre ces deux plantes et de l'absence d'éléments similaires dans des génomes qui leur sont proches, il est probable que la présence de ce rétrotransposon chez la pomme de terre sauvage résulte d'un transfert horizontal entre les deux espèces hôtes plutôt que de l'évolution de la famille *Tvv1* à partir d'un ancêtre rétroélément commun.

La publication de la séquence complète du génome de l'accession PN40024 offre désormais la possibilité de caractériser la totalité des éléments transposables de la vigne. L'ensemble des données qui ont été accumulées au cours de la thèse contribue à compléter l'annotation du génome et pourra aider à perfectionner les programmes informatiques de recherche systématique des séquences répétées et des éléments transposables de la vigne.

Analyse transcriptionnelle

L'analyse transcriptionnelle de dix-huit familles de rétrotransposons a été réalisée par plusieurs approches complémentaires.

La recherche, dans les banques de données, de séquences EST correspondant à ces familles nous a permis d'obtenir des premières preuves de la transcription de rétroéléments chez la vigne. Ainsi, certains rétrotransposons s'expriment dans des variétés de vigne différentes, notamment dans le Chardonnay et le Cabernet sauvignon, et dans des conditions particulières, comme en réponse à un stress abiotique. Compte tenu du fait que des familles encore inconnues sont susceptibles de s'exprimer chez la vigne, le nombre de séquences EST de

rétrotransposons détecté dans cette étude semble indiquer que les rétroéléments sont peut-être transcrits plus fortement chez la vigne que chez d'autres plantes cultivées. Un moyen de répondre à cette question serait d'élargir nos critères de recherche pour détecter les séquences EST correspondant à d'autres éléments transposables, en nous appuyant, par exemple, sur les séquences d'éléments déjà décrits ou en utilisant des motifs conservés entre les différentes familles d'éléments.

L'évaluation du niveau de transcription de seize familles de rétrotransposons par RT-PCR semi-quantitative nous a permis de mettre en évidence que la présence de leurs transcrits est régulée dans les organes végétatifs et reproducteurs de la vigne en conditions normales de croissance. Ainsi, la majorité des familles étudiées présente un profil d'expression spécifique à chaque type de tissu. Cependant, la famille *Noble*, qui contient certaines copies potentiellement autonomes, est transcrite dans la plupart des organes que nous avons étudiés, ce qui peut résulter de la présence de copies de cette famille dans un environnement épigénétique favorable.

L'expression des rétrotransposons est généralement plus forte dans les méristèmes apicaux et dans les embryons somatiques que dans les autres tissus, et pourrait participer à l'apparition de variabilité somaclonale.

L'évaluation de l'effet de deux stress sur les niveaux de transcription des seize familles dans les feuilles de vigne a révélé une augmentation de la quantité de leurs transcrits en réponse à la blessure et à l'infection par le mildiou. Ces résultats suggèrent que les nombreux stress subis par la vigne au cours de sa longue période de culture peuvent avoir influencé la mobilité des rétrotransposons.

La régulation de la transcription est une étape importante du contrôle de l'activité des rétrotransposons et l'insertion de nouvelles copies suite à l'induction de la transcription par les stress a déjà été décrite chez plusieurs espèces végétales. Aussi, il est très probable que les stress aient les mêmes conséquences chez la vigne et que les éléments mobiles aient participé à l'apparition de diversité génétique et à la différenciation des clones de vigne. La stratégie S-SAP, qui offre un complément intéressant à l'étude du niveau de transcription puisqu'elle permet de détecter de nouvelles insertions, pourrait être utilisée pour évaluer la mobilité des éléments transposables dans le génome de clones de vigne stressés. Cependant, les cas de mutations somatiques affectant la totalité des cellules d'un clone sont rares et la présence de

nouvelles insertions dans quelques cellules ou dans une seule assise cellulaire (chimères) est difficile à détecter.

La caractérisation des séquences adjacentes aux sites d'insertions des éléments transposables dans le génome de l'accession PN40024 pourrait nous permettre d'identifier des copies susceptibles d'avoir participé à la différenciation des variétés de vigne en ayant perturbé ou modifié l'expression de gènes d'intérêt agronomique, comme des gènes de résistance aux pathogènes ou de la teneur des baies en composés aromatiques.

L'identification de transcrits chimériques nous permet de supposer que les rétrotransposons peuvent avoir, au delà des effets directement provoqués par leur mobilité, des effets sur l'expression de gènes de la vigne, notamment en participant à la modification de leurs profils d'expression ou en partageant avec les gènes adjacents la régulation épigénétique qui les cible.

La diversité de séquence des ADNc correspondant aux transcrits de rétrotransposons exprimés dans les feuilles blessées montre que plusieurs copies de chaque famille s'expriment. Dans le cas de la famille *Tvv1*, les séquences UTL des copies exprimées sont très variables, ce qui devrait nous permettre d'identifier précisément les copies matrices, d'étudier plus particulièrement leur profil d'expression et de caractériser les motifs LTR susceptibles de diriger leur expression.

Création de nouveaux marqueurs moléculaires

De part leur nature et leur distribution dans les génomes, les rétrotransposons sont potentiellement des marqueurs moléculaires très polymorphes. Les techniques de REMAP (Retro Element Microsatellite Amplified Polymorphism) et IRAP (Inter Retroelement Amplified Polymorphism), qui s'appuient sur les domaines LTRs des rétrotransposons, se sont révélées efficaces pour le développement de marqueurs moléculaires chez les plantes (Kumar and Hirochika, 2001; Schulman *et al.*, 2004). Ces deux techniques ont notamment permis la cartographie de gènes (Manninen *et al.*, 2000) et l'étude de la stabilité génomique dans des espèces allopolyploïdes (Baumel *et al.*, 2002), la distinction d'espèces au sein d'un même genre (Kalendar *et al.*, 1999) et l'évaluation de la diversité génétique de plantes multipliées végétativement (Breto *et al.*, 2001).

Chez la vigne, l'un des objectifs à court terme est la recherche de nouveaux marqueurs moléculaires permettant d'évaluer la diversité dans les collections de vignes, d'aider à la prospection de nouveaux clones et de protéger les clones agréés. A partir d'amorces dessinées

dans l'élément *Gret1*, des marqueurs moléculaires REMAP et IRAP ont permis de différencier les variétés, mais pas les clones à l'intérieur d'un même cépage (Pereira *et al.*, 2005). Les expériences de S-SAP que nous avons réalisées au laboratoire n'ont porté jusqu'à présent que sur les trois premiers éléments décrits, *Gret1*, *Tvv1* et *Vine-1*, sans nous permettre de montrer de manière reproductible du polymorphisme inter-clonal.

La caractérisation de nouvelles familles de rétrotransposons du génome de la vigne, notamment celles qui, comme les familles *Noble* et *Edel*, présentent un grand nombre de copies et une activité transcriptionnelle importante, permettra peut-être de mettre en évidence du polymorphisme entre les clones et de développer de nouveaux marqueurs moléculaires. Certains éléments, comme *Tvv1* et *Noble*, contiennent une région UTL très variable qui permet d'établir, par une seule amplification, des profils multibandes spécifiques à chaque génotype. Cette technique, baptisée RUP, a déjà été utilisée avec *Tvv1* pour différencier les variétés mais ne permet pas de distinguer les clones les uns des autres. En développant ce marquage grâce à l'amplification de la région UTL de l'élément *Noble*, il nous sera peut être possible de parvenir à ce niveau de distinction.

La sélection et la multiplication par l'homme des clones présentant des caractères d'intérêt agronomique a pu contribuer au maintien d'éléments transposables qui auraient été naturellement contre-sélectionnés. L'induction de l'expression des rétrotransposons par les stress permet d'imaginer que ces éléments ont pu jouer un rôle dans l'évolution du génome de la vigne en lui apportant, en réponse aux variations climatiques, aux interactions avec d'autres organismes et aux interventions de l'homme, une partie de la plasticité génétique nécessaire à son adaptation et à son évolution. Ainsi, les éléments transposables contribuent peut-être à la richesse du patrimoine viticole et à la production de vins si différents à travers le monde.

Bibliographie

Bibliographie

- Adam-Blondon, A. F., C. Roux, D. Claux, G. Butterlin, D. Merdinoglu and P. This (2004). "Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics." *Theor Appl Genet* 109(5): 1017-27.
- Alleweldt, G. and J. V. Possingham (1988). "Progress in grapevine breeding." *Theoretical and Applied Genetics* 75: 699_673.
- Antonius-Klemola, K., R. Kalendar and A. H. Schulman (2006). "TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports." *Theor Appl Genet* 112(6): 999-1008.
- Anxolabehere, D., M. G. Kidwell and G. Periquet (1988). "Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile elements." *Molecular Biology and Evolution* 5: 252-269.
- Aradhya, M. K., G. S. Dangl, B. H. Prins, J. M. Boursiquot, M. A. Walker, C. P. Meredith and C. J. Simon (2003). "Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L." *Genet Res* 81(3): 179-92.
- Avery, O. T., C. M. McLeod and M. McCarthy (1944). "Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III." *J. Exp. Med* 79: 137-158.
- Baumel, A., M. Ainouche, R. Kalendar and A. H. Schulman (2002). "Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae)." *Mol Biol Evol* 19(8): 1218-27.
- Bell, A. C., A. G. West and G. Felsenfeld (2001). "Insulators and boundaries: versatile regulatory elements in the eukaryotic." *Science* 291(5503): 447-50.
- Bennetzen, J. L., P. SanMiguel, C. MingSheng., A. Tikhonov, M. Francki and Z. Avramova (1998). "Grass genomes." *Proceedings of the National Academy of Sciences of the USA* 95(5): 1975-1978.
- Biemont, C., A. Tsitrone, C. Vieira and C. Hoogland (1997). "Transposable element distribution in *Drosophila*." *Genetics* 147(4): 1997-9.
- Biemont, C. and C. Vieira (2005). "What transposable elements tell us about genome organization and evolution: the case of *Drosophila*." *Cytogenet Genome Res* 110(1-4): 25-34.
- Billiard, R. (1913). *La vigne dans l'Antiquité*. Lyon.
- Bortoluzzi, S., F. d'Alessi, C. Romualdi and G. A. Danieli (2000). "The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach." *Genome Res* 10(3): 344-9.
- Boss, P. K. and M. R. Thomas (2002). "Association of dwarfism and floral induction with a grape 'green revolution' mutation." *Nature* 416(6883): 847-850.
- Boursiquot, J. M. and P. This (1999). "Essai de définition du cépage." *Progrès Agricole et Viticole* 116: 359-361.
- Boutabout, M., M. Wilhelm and F. X. Wilhelm (2001). "DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1." *Nucleic Acids Research* 29(11): 2217-2222.
- Bowen, N. and J. McDonald (1999). "Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements." *Genome Research* 9: 924-35.
- Bowers, J., J. M. Boursiquot, P. This, K. Chu, H. Johansson and C. Meredith (1999). "Historical Genetics: The Parentage of Chardonnay, Gamay, and Other Wine Grapes of Northeastern France." *Science* 285(5433): 1562-1565.

- Bowers, J. E. and C. P. Meredith (1997). "The parentage of a classic wine grape, Cabernet Sauvignon." *Nat Genet* 16(1): 84-7.
- Breto, M. P., C. Ruiz, J. A. Pina and M. J. Asins (2001). "The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species." *Molecular Phylogenetics and Evolution* 21: 285-293.
- Briggs, J. A., M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt and M. C. Johnson (2004). "The stoichiometry of Gag protein in HIV-1." *Nat Struct Mol Biol* 11(7): 672-5.
- Britten, R. J. and E. H. Davidson (1969). "Gene regulation for higher cells: a theory." *Science* 165: 349-357.
- Bronner, A. and J. Oliveira (1990). "Creation and study of the Pinot Noir variety lineage." *Proc. 5th Intern. Symp. Grape Breeding. St Martin/Pflaz, Germany, Sept 1989, Vitis*(special issue): 69-80.
- Brosius, J. (1991). "Retroposons-seeds of evolution." *Science* 251(4995): 753.
- Brunner, A. M., V. B. Busov and S. H. Strauss (2004). "Poplar genome sequence: functional genomics in an ecologically dominant plant species." *Trends Plant Sci* 9(1): 49-56.
- Bureau, T. E. and S. R. Wessler (1994). "*Tourist* : a large family of small inverted repeat elements frequently associated with maize genes." *Plant cell* 4(10): 1283-1294.
- Bureau, T. E., P. C. Ronald and S. R. Wessler (1996). "A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes." *Proc Natl Acad Sci U S A* 93(16): 8524-8529.
- Cai, H. N. and M. Levine (1997). "The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo." *Embo J* 16(7): 1732-41.
- Camirand, A. and N. Brisson (1990). "The complete nucleotide sequence of the Tst1 retrotransposon of potato." *Nucleic Acids Research* 18: 4929.
- Carbonneau, A., A. Deloire and B. Jaillard (2007). *La vigne : physiologie, terroir, culture*, DUNOD.
- Casacuberta, E. and M. L. Pardue (2005). "HeT-A and TART, two *Drosophila* retrotransposons with a bona fide role in chromosome structure for more than 60 million years." *Cytogenet Genome Res* 110(1-4): 152-9.
- Casacuberta, E. and M. L. Pardue (2006). "RNA interference has a role in regulating *Drosophila* telomeres." *Genome Biol* 7(5): 220.
- Casacuberta, J. M., S. Vernhettes and M.-A. Grandbastien (1995). "Sequence variability within the tobacco retrotransposon Tnt1 population." *EMBO Journal* 14(11): 2670-2678.
- Casacuberta, J. M., S. Vernhettes, C. Audeon and M.-A. Grandbastien (1997). "Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution." *Genetica* 100: 109-117.
- Chandler, V. L. and K. J. Hardeman (1992). "The Mu elements of *Zea mays*." *Adv Genet* 30: 77-122.
- Chargaff, E. (1950). "Chemical specificity of nucleic acids and mechanism of their enzymatic degradation." *Experientia* 6: 201.
- Charlesworth, B., P. Sniegowski and W. Stephan (1994). "The evolutionary dynamics of repetitive DNA in eukaryotes." *Nature* 371(6494): 215-20.
- Coen, E. S., R. Carpenter and C. Martin (1986). "Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*." *Cell* 47(2): 285-96.
- Cohen, D. E., L. S. Davidow, J. A. Erwin, N. Xu, D. Warshawsky and J. T. Lee (2007). "The DXPas34 repeat regulates random and imprinted X inactivation." *Dev Cell* 12(1): 57-71.

- Comfort, N. C. (2001). "From controlling elements to transposons : Barbara McClintock and the Nobel Prize." *Trends in Genetics* 17(8): 475-478.
- Crick, F. H. C., J. S. Griffith and L. E. Orgel (1957). "Codes without commas." *Proc Natl Acad Sci U S A* 43: 416-421.
- Curcio, M. J. and D. J. Garfinkel (1999). "New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIF." *Trends in Genetics* 15(2): 43-45.
- Dalton, K., R. Casais, K. Shaw, K. Stirrups, S. Evans, T. D. Brown, P. Britton and D. Cavanagh (2001). "Sequences required for replication and packaging of IBV RNA." *Adv Exp Med Biol* 494: 553-6.
- Daniels, S., K. Peterson, L. Strausbaugh, M. Kidwell and A. Chovnick (1990). "Evidence for horizontal transmission of the P transposable element between *Drosophila* species." *Genetics* 124(2): 339-355.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford.
- de Araujo, P. G., J. M. Casacuberta, A. P. Costa, R. Y. Hashimoto, M. A. Grandbastien and M. A. Van Sluys (2001). "Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus." *Mol Genet Genomics* 266(1): 35-41.
- de Araujo, P. G., M. Rossi, E. M. de Jesus, N. L. Saccaro, Jr., D. Kajihara, R. Massa, J. M. de Felix, R. D. Drummond, M. C. Falco, S. M. Chabregas, E. C. Ulian, M. Menossi and M. A. Van Sluys (2005). "Transcriptionally active transposable elements in recent hybrid sugarcane." *Plant J* 44(5): 707-17.
- Dellaporta, S. L., P. S. Chomet, J. P. Mottinger, J. A. Wood, S. M. Yu and J. B. Hicks (1984). "Endogenous transposable elements associated with virus infection in maize." *Cold Spring Harb Symp Quant Biol* 49: 321-8.
- Deragon, J. M., N. Gilbert, L. Rouquet, A. Lenoir, P. Arnaud and G. Picard (1996). "A transcriptional analysis of the S1Bn (*Brassica napus*) family of SINE retroposons." *Plant Mol Biol* 32(5): 869-78.
- Devos, K. M., J. K. Brown and J. L. Bennetzen (2002). "Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*." *Genome Res* 12(7): 1075-9.
- Diao, X., M. Freeling and D. Lisch (2006). "Horizontal Transfer of a Plant Transposon." *PLoS Biology* 4(1): e5.
- Diez-Navajas, A., C. Greif, A. Poutaraud and D. Merdinoglu (2006). "Two simplified fluorescent staining techniques to observe infection structures of the oomycete *Plasmopara viticola* in grapevine leaf tissues." *Micron*: doi:10.1016/j.micron.2006.09.009.
- Dixit, A., K. H. Ma, J. W. Yu, E. G. Cho and Y. J. Park (2006). "Reverse transcriptase domain sequences from Mungbean (*Vigna radiata*) LTR retrotransposons: sequence characterization and phylogenetic analysis." *Plant Cell Rep* 25(2): 100-11.
- Doligez, A., A. F. Adam-Blondon, G. Cipriani, G. Di Gaspero, V. Laucou, D. Merdinoglu, C. P. Meredith, S. Riaz, C. Roux and P. This (2006). "An integrated SSR map of grapevine based on five mapping populations." *Theor Appl Genet* 113(3): 369-82.
- Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." *Nature* 284(5757): 601-603.
- Dover, G. and W. F. Doolittle (1980). "Modes of genome evolution." *Nature* 288(5792): 646-7.
- Drake, J. W., B. Charlesworth, D. Charlesworth and J. F. Crow (1998). "Rates of spontaneous mutation." *Genetics* 148: 1667-1686.
- Ebner, P. D. and D. J. O'Callaghan (2006). "Genetic complexity of EHV-1 defective interfering particles and identification of novel IR4/UL5 hybrid proteins produced during persistent infection." *Virus Genes* 32(3): 313-20.

- Evans, M. J., E. Bacharach and S. P. Goff (2004). "RNA sequences in the Moloney murine leukemia virus genome bound by the Gag precursor protein in the yeast three-hybrid system." *J Virol* 78(14): 7677-84.
- Feschotte, C., N. Jiang and S. R. Wessler (2002). "Plant transposable elements: Where genetics meets genomics." *Nature Reviews Genetics* 3(5): 329-341.
- Flavell, R. B. (1986). "Repetitive DNA and chromosome evolution in plants." *Philos Trans R Soc Lond B Biol Sci* 312(1154): 227-42.
- Gabriel, A., M. Willems, E. H. Mules and J. D. Boeke (1996). "Replication infidelity during a single cycle of Ty1 retrotransposition." *Proc Natl Acad Sci U S A* 93(15): 7767-71.
- Galet, P. (1993). *Précis de viticulture*, Imprimerie Déhan, Montpellier.
- Gamow, G. (1954). "Possible relation between deoxyribonucleic acid and protein structures." *Nature*: 173-318.
- Gao, L., E. M. McCarthy, E. W. Ganko and J. F. McDonald (2004). "Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences." *BMC Genomics* 5(1): 18.
- Gaut, B. S., B. R. Morton, B. C. Mccaig and M. T. Clegg (1996). "Substitution rate comparisons between grasses and palms : synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*." *Proc Natl Acad Sci U S A* 93: 10274-10279.
- Gerasimova, T. I., K. Byrd and V. G. Corces (2000). "A chromatin insulator determines the nuclear localization of DNA." *Mol Cell* 6(5): 1025-35.
- Gierl, A., H. Saedler and P. A. Peterson (1989). "Maize transposable elements." *Annu Rev Genet* 23: 71-85.
- Gierl, A. (1996). "The *En/Spm* transposable element of maize." *Cur Top Microbiol Immunol* 204: 145-159.
- Grandbastien, M.-A., A. Spielmann and M. Caboche (1989). "*Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics." *Nature* 337(6205): 376-380.
- Grandbastien, M.-A., H. Lucas, J.-B. Morel, C. Mhiri, S. Vernhettes and J. Casacuberta (1997). "The expression of the tobacco *Tnt1* retrotransposon is linked to plant defense responses." *Genetica* 1-3: 241-252.
- Grandbastien, M.-A. (1998). "Activation of plant retrotransposons under stress conditions." *Trends in Plant Science* 3(5): 181-187.
- Griffith, F. (1928). "The significance of pneumococcal types." *J. Hyg.* 27: 13-159.
- Hartmann, H. T. and D. E. Kester (1975). *Plant propagation*. New Jersey.
- Hashida, S. N., K. Kitamura, T. Mikami and Y. Kishima (2003). "Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in *Antirrhinum majus*." *Plant Physiol* 132(3): 1207-16.
- Havecker, E. R., X. Gao and D. F. Voytas (2004). "The diversity of LTR retrotransposons." *Genome Biol* 5(6): 225.
- He, Z. H., H. T. Dong, J. X. Dong, D. B. Li and P. C. Ronald (2000). "The rice *Rim2* transcript accumulates in response to *Magnaporthe grisea* and its predicted protein product shares similarity with TNP2-like proteins encoded by CACTA transposons." *Mol Gen Genet* 264(1-2): 2-10.
- Hershey, A. D. (1946). "Spontaneous mutations in bacterial viruses." *Cold Spring Harbor Symposia on Quantitative Biology* XI: 33_37.
- Hirochika, H. (1993). "Activation of tobacco retrotransposons during tissue culture." *Embo J* 12(6): 2521-8.

- Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa and M. Kanda (1996). "Retrotransposons of rice involved in mutations induced by tissue culture." *Proc Natl Acad Sci U S A* 93(15): 7783-8.
- Hocquigny, S., F. Pelsy, V. Dumas, S. Kindt, M. C. Heloir and D. Merdinoglu (2004). "Diversification within grapevine cultivars goes through chimeric states." *Genome* 47(3): 579-89.
- Hu, W., O. P. Das and J. Messing (1995). "Zeon-1, a member of a new maize retrotransposon family." *Molecular & General Genetics: MGG* 248(4): 471-480.
- International Rice Genome Sequencing Project (2005). "The map-based sequence of the rice genome." *Nature* 436(7052): 793-800.
- Jaaskelainen, M., A. H. Mykkanen, T. Arna, C. M. Vicent, A. Suoniemi, R. Kalendar, H. Savilahti and A. H. Schulman (1999). "Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells." *Plant Journal* 20(4): 413-422.
- Jacob, F. and J. Monod (1965). "Genetic mapping of the elements of the lactose region in *Escherichia coli*." *Biochemical and Biophysical Research Communications* 18(5-6): 693_701.
- Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Huguency, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyere, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pe, G. Valle, M. Morgante, M. Caboche, A. F. Adam-Blondon, J. Weissenbach, F. Quetier and P. Wincker (2007). "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla." *Nature* 449(7161): 463-7.
- Jiang, J., J. A. Birchler, W. A. Parrott and R. K. Dawe (2003a). "A molecular view of plant centromeres." *Trends Plant Sci* 8(12): 570-5.
- Jiang, N., Z. Bao, S. Temnykh, Z. Cheng, J. Jiang, R. A. Wing, S. R. McCouch and S. R. Wessler (2002a). "Dasheng: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice." *Genetics* 161(3): 1293-305.
- Jiang, N., I. K. Jordan and S. R. Wessler (2002b). "Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome." *Plant Physiol* 130(4): 1697-705.
- Jiang, N., Z. Bao, X. Zhang, H. Hirochika, S. R. Eddy, S. R. McCouch and S. R. Wessler (2003b). "An active DNA transposon family in rice." *Nature* 421: 163-167.
- Jin, Y. K. and J. L. Bennetzen (1989). "Structure and coding properties of Bs1, a maize retrovirus-like transposon." *Proc Natl Acad Sci U S A* 86(16): 6235-9.
- Johannsen, W. L. (1909). *Elemente der exakten Erblichkeitslehre*. Jena.
- Johns, M. A., J. Mottinger and M. Freeling (1985). "A low copy number, copia-like transposon in maize." *Embo J* 4(5): 1093-101.
- Kajikawa, M. and N. Okada (2002). "LINEs mobilize SINEs in the eel through a shared 3' sequence." *Cell* 111(3): 433-444.
- Kalendar, R., T. Grob, M. Regina, A. Suoniemi and A. Schulman (1999). "IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques." *Theoretical and Applied Genetics* 98(5): 704-711.

- Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo and A. H. Schulman (2000). "Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence." *Proceedings of the National Academy of Sciences of the USA* 97: 6603-6607.
- Kalendar, R., C. M. Vicent, O. Peleg, K. Ananthawat-Jonsson, A. Bolshoy and A. H. Schulman (2004). "Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes." *Genetics* 166(3): 1437-50.
- Kalmykova, A. I., D. A. Kwon, Y. M. Rozovsky, N. Hueber, P. Capy, C. Maisonhaute and V. A. Gvozdev (2004). "Selective expansion of the newly evolved genomic variants of retrotransposon 1731 in the *Drosophila* genomes." *Mol Biol Evol* 21(12): 2281-9.
- Kapitonov, V. V. and J. Jurka (2001). "Rolling-circle transposons in eukaryotes." *Proc Natl Acad Sci U S A* 98(15): 8714-8719.
- Kashkush, K., M. Feldman and A. A. Levy (2003). "Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat." *Nat Genet* 33(1): 102-6.
- Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." *Science* 303(5664): 1626-32.
- Kempken, F. and F. Windhofer (2001). "The *hAT* family : a versatile transposon group common to plants, fungi, animals, and man." *Chromosoma* 110(1): 1-9.
- Kidwell, M. G. and D. R. Lisch (2001). "Perspective : transposable elements, parasitic DNA, and genome evolution." *Evolution* 55(1): 1-24.
- Kidwell, M. G. (2002). "Transposable elements and the evolution of genome size in eukaryotes." *Genetica* 115(1): 49-63.
- Kikuchi, K., K. Terauchi, M. Wada and H. Hirano (2003). "The plant MITE mPing is mobilized in anther culture." *Nature* 421: 167-170.
- Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prud'homme and A. Bucheton (1994). "Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*." *Proc Natl Acad Sci U S A* 91(4): 1285-9.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge.
- Kitamura, K., S. N. Hashida, T. Mikami and Y. Kishima (2001). "Position effect of the excision frequency of the *Antirrhinum* transposon Tam3: implications for the degree of position-dependent methylation in the ends of the element." *Plant Mol Biol* 47(4): 475-90.
- Klekowski, E. J., Jr. (1998). "Mutation rates in mangroves and other plants." *Genetica* 102/103: 325-331.
- Kobayashi, S., N. Goto-Yamamoto and H. Hirochika (2004). "Retrotransposon-induced mutations in grape skin color." *Science* 304: 982.
- Konieczny, A., D. F. Voytas, M. P. Cumming and F. M. Ausubel (1991). "A superfamily of *Arabidopsis thaliana* retrotransposons." *Genetics* 127: 801-809.
- Kotnova, A. P., I. A. Glukhov, N. N. Karpova, V. B. Salenko, N. V. Lyubomirskaya and Y. V. Ilyin (2007). "Evidence for recent horizontal transfer of gypsy-homologous LTR-retrotransposon gtwin into *Drosophila erecta* followed by its amplification with multiple aberrations." *Gene* 396(1): 39-45.
- Kumar, A. and H. Hirochika (2001). "Applications of retrotransposons as genetic tools in plant biology." *Trends Plant Sci* 6(3): 127-34.
- Kumar, S., K. Tamura and M. Nei (2004). "MEGA3 : Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment." *Briefings in Bioinformatics* 5: 150-163.

- Labra, M., S. Imazio, F. Grassi, M. Rossoni and F. Sala (2004). "*Vine-1* retrotransposon-based sequence-specific amplified polymorphism for *Vitis vinifera* L. genotyping." *Plant Breeding Germany* 123(2): 180-185.
- Lamotte, O., K. Gould, D. Lecourieux, A. Sequeira-Legrand, A. Lebrun-Garcia, J. Durner, A. Pugin and D. Wendehenne (2004). "Analysis of nitric oxide signaling functions in tobacco cells challenged by the elicitor cryptogein." *Plant Physiol* 135(1): 516-29.
- Lankenau, D. H., P. Huijser, E. Jansen, K. Miedema and W. Hennig (1990). "DNA sequence comparison of micropia transposable elements from *Drosophila hydei* and *Drosophila melanogaster*." *Chromosoma* 99(2): 111-7.
- Laten, H. M., E. R. Havecker, L. M. Farmer and D. F. Voytas (2003). "SIRE1, an endogenous retrovirus family from *Glycine max*, is highly homogeneous and evolutionarily young." *Mol Biol Evol* 20(8): 1222-30.
- Lazzarini, R. A., J. D. Keene and M. Schubert (1981). "The origins of defective interfering particles of the negative-strand RNA viruses." *Cell* 26(2 Pt 2): 145-54.
- Le Grice, S. F. (2003). "'In the beginning': initiation of minus strand DNA synthesis in retroviruses and LTR-containing retrotransposons." *Biochemistry* 42(49): 14349-55.
- Lee, D. L., T. H. Ellis, L. Turner, R. P. Hellens and W. G. Cleary (1990). "A copia-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis." *Plant Mol Biol* 15(5): 707-22.
- Levadoux, L. (1956). "Les populations sauvages et cultivées de *Vitis vinifera* L." *Annales de l'Amélioration des Plantes* 6: 59-118.
- Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar and F. M. Sheen (1993). "Transposons in place of telomeric repeats at a *Drosophila* telomere." *Cell* 75(6): 1083-93.
- Li, W., P. Zhang, J. P. Fellers, B. Friebe and B. S. Gill (2004). "Sequence composition, organization, and evolution of the core Triticeae genome." *Plant J* 40(4): 500-11.
- Liu, J., Y. He, R. Amasino and X. Chen (2004). "siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*." *Genes Dev* 18(23): 2873-8.
- Lyon, M. F. (2006). "Do LINEs Have a Role in X-Chromosome Inactivation?" *J Biomed Biotechnol* 2006(1): 59746.
- Madlung, A. and L. Comai (2004). "The effect of stress on genome regulation and structure." *Annals of Botany* 94(4): 481-495.
- Manetti, M. E., M. Rossi, A. P. Costa, A. M. Clausen and M. A. Van Sluys (2007). "Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera." *BMC Evol Biol* 7: 34.
- Manninen, I. and A. H. Schulman (1993). "BARE-1, a copia-like retroelement in barley (*Hordeum vulgare* L.)." *Plant Molecular Biology* 22(5): 829-846.
- Manninen, O., R. Kalendar, J. Robinson and A. H. Schulman (2000). "Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley." *Mol Gen Genet* 264(3): 325-34.
- Marillonnet, S. and S. R. Wessler (1998). "Extreme structural heterogeneity among the members of a maize retrotransposon family." *Genetics* 150(3): 1245-1256.
- Martienssen, R. and A. Baron (1994). "Coordinate suppression of mutations caused by Robertson's mutator transposons in maize." *Genetics* 136(3): 1157-70.
- Masson, P., R. Surosky, J. A. Kingsbury and N. V. Fedoroff (1987). "Genetic and molecular analysis of the Spm-dependent a-m2 alleles of the maize a locus." *Genetics* 117(1): 117-37.
- McClintock, B. (1941). "The Association of Mutants with Homozygous Deficiencies in *Zea Mays*." *Genetics* 26(5): 542-71.

- McClintock, B. (1942). "The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion." *Proc Natl Acad Sci U S A* 28(11): 458-63.
- McClintock, B. (1944). "The Relation of Homozygous Deficiencies to Mutations and Allelic Series in Maize." *Genetics* 29(5): 478-502.
- McClintock, B. (1956a). "Controlling elements and the gene." *Cold Spring Harb Symp Quant Biol* 21: 197-216.
- McClintock, B. (1956b). "Intranuclear systems controlling gene action and mutation." *Brookhaven Symp Biol*(8): 58-74.
- McClintock, B. (1961). "Some parallels between gene control systems in maize and in bacteria." *Am. Nat.* 95: 265-277.
- McClintock, B. (1984). "The significance of responses of the genome to challenge." *Science* 226: 792-801.
- McClure, M. A., M. S. Johnson, D. F. Feng and R. F. Doolittle (1988). "Sequence Comparisons of Retroviral Proteins: Relative Rates of Change and General Phylogeny." *Proceedings of the National Academy of Sciences of the United States of America* 85: 2469-2473.
- Meignin, C., J. L. Bailly, F. Arnaud, B. Dastugue and C. Vaury (2003). "The 5' untranslated region and Gag product of Idefix, a long terminal repeat-retrotransposon from *Drosophila melanogaster*, act together to initiate a switch between translated and untranslated states of the genomic mRNA." *Mol Cell Biol* 23(22): 8246-54.
- Melayah, D., E. Bonnard, B. Chalhoub, C. Audeon and M. A. Grandbastien (2001). "The mobility of the tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors." *The Plant Journal* 28(2): 159-168.
- Meyers, B. C., S. V. Tingey and M. Morgante (2001). "Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome." *Genome Research* 11(10): 1660-1676.
- Moisy, C., S. Blanc, D. Merdinoglu and F. Pelsy (2008). "Structural variability of Tvv1 grapevine retrotransposons can be caused by illegitimate recombination." *Theor Appl Genet.*
- Moncada, X., F. Pelsy, D. Merdinoglu and P. Hinrichsen (2006). "Genetic diversity and geographical dispersal in grapevine clones revealed by microsatellite markers." *Genome* 49(11): 1459-72.
- Morgan, H. D., H. G. Sutherland, D. I. Martin and E. Whitelaw (1999). "Epigenetic inheritance at the agouti locus in the mouse." *Nat Genet* 23(3): 314-8.
- Morgan, T. H. (1910). "Sex-limited inheritance in *Drosophila*." *Science* 32: 120-122.
- Moser, C., C. Segala, P. Fontana, I. Salakhudtinov, P. Gatto, M. Pindo, E. Zyprian, R. Toepfer, M. S. Grandó and R. Velasco (2005). "Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera* L." *Funct. Integr. Genomics* 5(4): 2008-217.
- Nekrutenko, A. and W. H. Li (2001). "Transposable elements are found in a large number of human protein-coding genes." *Trends in Genetics* 17: 619-621.
- Neumann, P., D. Pozarkova and J. Macas (2003). "Highly abundant pea LTR retrotransposon Ogré is constitutively transcribed and partially spliced." *Plant Mol Biol* 53(3): 399-410.
- Nigumann, P., K. Redik, K. Matlik and M. Speek (2002). "Many human genes are transcribed from the antisense promoter of L1 retrotransposon." *Genomics* 79(5): 628-34.
- Ohtsubo, H., N. Kume-kawa and E. Ohtsubo (1999). "RIRE2, a novel gypsy-type retrotransposon from rice." *Genes & Genetic Systems* 74(3): 83-91.
- Okamoto, H. and H. Hirochika (2001). "Silencing of transposable elements in plants." *Trends Plant Sci* 6(11): 527-34.

- Orgel, L. E. and F. H. Crick (1980). "Selfish DNA: the ultimate parasite." *Nature* 284(5757): 604-607.
- Orgel, L. E., F. H. Crick and C. Sapienza (1980). "Selfish DNA." *Nature* 288(5792): 645-6.
- Paillart, J. C., M. Shehu-Xhilaga, R. Marquet and J. Mak (2004). "Dimerization of retroviral RNA genomes: an inseparable pair." *Nat Rev Microbiol* 2(6): 461-72.
- Pardue, M. L., S. Rashkova, E. Casacuberta, P. G. DeBaryshe, J. A. George and K. L. Traverse (2005). "Two retrotransposons maintain telomeres in *Drosophila*." *Chromosome Res* 13(5): 443-53.
- Pearce, S. R., C. Stuart-Rogers, M. R. Knox, A. Kumar, T. H. Ellis and A. J. Flavell (1999). "Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies." *Plant J* 19(6): 711-7.
- Pelisson, A., L. Mejlumian, V. Robert, C. Terzian and A. Bucheton (2002). "*Drosophila* germline invasion by the endogenous retrovirus gypsy: involvement of the viral env gene." *Insect Biochem Mol Biol* 32(10): 1249-56.
- Pelsy, F. and D. Merdinoglu (2002). "Complete sequence of Tv1, a family of Ty 1 copia-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking." *Theor Appl Genet* 105(4): 614-621.
- Pelsy, F. (2007). "Untranslated leader region polymorphism of Tv1, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*." *Theor Appl Genet* 116(1): 15-27.
- Peng, F., K. E. Reid, N. Liao, J. Schlosser, D. Lijavetzky, R. Holt, J. M. Martinez Zapater, S. Jones, M. Marra, J. Bohlmann and S. T. Lund (2007). "Generation of ESTs in *Vitis vinifera* wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and discovery of new candidate genes with potential roles in berry development." *Gene* 402(1-2): 40-50.
- Pereira, H. S., A. Barao, M. Delgado, L. Morais-Cecilio and W. Viegas (2005). "Genomic analysis of Grapevine Retrotransposon 1 (Gret 1) in *Vitis vinifera*." *Theor Appl Genet* 111(5): 871-8.
- Perrin, M., C. Gertz and J. E. Masson (2004). "High efficiency initiation of regenerable embryogenic callus from anther filaments of 19-grapevine genotypes grown worldwide." *Plant Science* 167: 1343-1349.
- Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D. S. Brar, S. Jackson, R. A. Wing and O. Panaud (2006). "Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice." *Genome Res* 16(10): 1262-9.
- Plasterk, R., Z. Izsvak and Z. Ivics (1999). "Resident aliens : the *Tc1/mariner* superfamily of transposable elements." *Trends in Genetics* 15(8): 326-332.
- Pouget, R. (1990). *Histoire de la lutte contre le phylloxéra de la vigne en France (1868-1895)*, INRA.
- Pouteau, S., E. Huttner, M. A. Grandbastien and M. Caboche (1991). "Specific expression of the tobacco Tnt1 retrotransposon in protoplasts." *Embo J* 10(7): 1911-8.
- Preston, B. D. (1996). "Error-prone retrotransposition: Rime of the ancient mutators." *PNAS* 93(15): 7427-7431.
- Riccioni, C., A. Rubini, B. Belfiori, V. Passeri, F. Paolocci and S. Arcioni (2008). "Tmt1: the first LTR-retrotransposon from a Tuber spp." *Curr Genet* 53(1): 23-34.
- Rico-Cabanas, L. and J. A. Martinez-Izquierdo (2007). "CIRE1, a novel transcriptionally active Ty1-copia retrotransposon from *Citrus sinensis*." *Mol Genet Genomics* 277(4): 365-77.
- Roetschi, A., A. Si-Ammour, L. Belbahri, F. Mauch and B. Mauch-Mani (2001). "Characterization of an Arabidopsis-Phytophthora pathosystem: resistance requires a

- functional PAD2 gene and is independent of salicylic acid, ethylene and jasmonic acid signalling." *Plant J* 28(3): 293-305.
- Rossi, M., P. G. Araujo and M. A. Van Sluys (2001). "Survey of transposable elements in sugarcane expressed sequence tags (EST)." *Genetics and Molecular Biology* 24(1-4): 147-154.
- Roulin, A., B. Piegu, R. A. Wing and O. Panaud (2007). "Evidence of multiple horizontal transfers of the LTR-retrotransposon RIRE1 within the genus *Oryza*." *Plant J*.
- Sabot, F., P. Sourdille and M. Bernard (2005). "Advent of a new retrotransposon structure: the long form of the Veju elements." *Genetica* 125(2-3): 325-32.
- Sabot, F. and A. H. Schulman (2006). "Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome." *Heredity* 97(6): 381-388.
- Sabot, F. and A. H. Schulman (2007). "Template switching can create complex LTR retrotransposon insertions in Triticeae genomes." *BMC Genomics* 8: 247.
- Sakai, H., T. Tanaka and T. Itoh (2007). "Birth and death of genes promoted by transposable elements in *Oryza sativa*." *Gene* 392: 59-63.
- Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular cloning : a laboratory manual*. NY., Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Sanchez-Garcia, A., X. Maside and B. Charlesworth (2005). "High rate of horizontal transfer of transposable elements in *Drosophila*." *Trends in Genetics* 21(4): 200-203.
- SanMiguel, P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P. S. Springer, K. J. Edwards, M. Lee, Z. Avramova and J. L. Bennetzen (1996). "Nested retrotransposons in the intergenic regions of the maize genome." *Science* 274(5288): 765-8.
- SanMiguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L. Bennetzen (1998). "The paleontology of intergene retrotransposons of maize." *Nat Genet* 20(1): 43-5.
- SanMiguel, P. J., W. Ramakrishna, J. L. Bennetzen, C. S. Busso and J. Dubcovsky (2002). "Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m)." *Funct Integr Genomics* 2(1-2): 70-80.
- Schueler, M. G. and B. A. Sullivan (2006). "Structural and functional dynamics of human centromeric chromatin." *Annu Rev Genomics Hum Genet* 7: 301-13.
- Schulman, A. H., A. J. Flavell and T. H. Ellis (2004). "The application of LTR retrotransposons as molecular markers in plants." *Methods Mol Biol* 260: 145-73.
- Sefc, K., F. Lefort, M. Grando, K. Scott, H. Steinkellner and M. R. Thomas (2001). "Microsatellite markers for grapevine: A state of the art." *Mol. Biol. Biotechnol Grapevine* 463: 433 - 463.
- Shamel, A. D. and C. S. Pomeroy (1936). "Bud mutations in horticultural crops." *Journal of Heredity*: 487-494.
- Shirasu, K., A. H. Schulman, T. Lahaye and P. Schulze-Lefert (2000). "A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion." *Genome Res* 10(7): 908-15.
- Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." *Nat Rev Genet* 8(4): 272-85.
- Sun, F. L., K. Haynes, C. L. Simpson, S. D. Lee, L. Collins, J. Wuller, J. C. Eissenberg and S. C. Elgin (2004). "cis-Acting determinants of heterochromatin formation on *Drosophila melanogaster* chromosome four." *Mol Cell Biol* 24(18): 8210-20.
- Sun, X., H. D. Le, J. M. Wahlstrom and G. H. Karpen (2003). "Sequence analysis of a functional *Drosophila* centromere." *Genome Res* 13(2): 182-94.
- Suoniemi, A., A. Narvanto and A. H. Schulman (1996). "The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays." *Plant Mol Biol* 31(2): 295-306.

- Takeda, S., K. Sugimoto, H. Otsuki and H. Hirochika (1998). "Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate." *Plant Molecular Biology* 36(3): 365-376.
- Takeda, S., K. Sugimoto, H. Otsuki and H. Hirochika (1999). "A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors." *Plant Journal* 18(4): 383-393.
- Temin, H. M. (1980). "Origin of retroviruses from cellular moveable genetic elements." *Cell* 21(3): 599-600.
- The Arabidopsis Initiative (2000). "Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*." *Nature* 408(6814): 796-815.
- The International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome." *Nature* 409(6822): 860-921.
- This, P., A. Jung, P. Boccacci, J. Borrego, R. Botta, L. Costantini, M. Crespan, G. S. Dangl, C. Eisenheld, F. Ferreira-Monteiro, S. Grando, J. Ibanez, T. Lacombe, V. Laucou, R. Magalhaes, C. P. Meredith, N. Milani, E. Peterlunger, F. Regner, L. Zulini and E. Maul (2004). "Development of a standard set of microsatellite reference alleles for identification of grape cultivars." *Theor Appl Genet* 109(7): 1448-58.
- Thomas, C. A., Jr. (1971). "The genetic organization of chromosomes." *Annu Rev Genet* 5: 237-56.
- Tomkins, J. P., D. G. Peterson, T. J. Yang, D. Main, E. F. Ablett, R. J. Henry, L. S. Lee, T. A. Holton, D. Waters and R. A. Wing (2001). "Grape (*Vitis vinifera* L.) BAC library construction, preliminary STC analysis, and identification of clones associated with flavonoid and stilbene biosynthesis." *American Journal of Enology and Viticulture* 52(4): 287-291.
- Turcich, M. P., A. Bokhari-Riza, D. A. Hamilton, C. He, W. Messier, C.-B. Stewart and J. P. Mascarenhas (1996). "PREM-2, a copia-type retroelement in maize is expressed preferentially in early microspores." *Sexual Plant Reproduction* 9(2): 65-74.
- Van de Lagemaat, L. N., J.-R. Landry, D. L. Mager and P. Medstrand (2003). "Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions." *Trends in Genetics* 19(10): 530-536.
- Van Loon, L. C. and E. A. Van Strien (1995). "The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins." *Physiological and Molecular Plant Pathology* 55: 85-97.
- Vernhettes, S., M. A. Grandbastien and J. M. Casacuberta (1998). "The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences." *Molecular Biology and Evolution* 15(7): 827-836.
- Verries, C., C. Bes, P. This and C. Tesniere (2000). "Cloning and characterization of Vine-1, a LTR-retrotransposon-like element in *Vitis vinifera* L., and other *Vitis* species." *Genome* 43(2): 366-376.
- Vettore, A. L., F. R. da Silva, E. L. Kemper, G. M. Souza, A. M. da Silva, M. I. Ferro, F. Henrique-Silva, E. A. Giglioti, M. V. Lemos, L. L. Coutinho, M. P. Nobrega, H. Carrer, S. C. Franca, M. Bacci Junior, M. H. Goldman, S. L. Gomes, L. R. Nunes, L. E. Camargo, W. J. Siqueira, M. A. Van Sluys, O. H. Thiemann, E. E. Kuramae, R. V. Santelli, C. L. Marino, M. L. Targon, J. A. Ferro, H. C. Silveira, D. C. Marini, E. G. Lemos, C. B. Monteiro-Vitorello, J. H. Tambor, D. M. Carraro, P. G. Roberto, V. G. Martins, G. H. Goldman, R. C. de Oliveira, D. Truffi, C. A. Colombo, M. Rossi, P. G. de Araujo, S. A. Sculaccio, A. Angella, M. M. Lima, V. E. de Rosa Junior, F. Siviero, V. E. Coscrato, M. A. Machado, L. Grivet, S. M. Di Mauro, F. G. Nobrega, C. F. Menck, M. D. Braga, G. P. Telles, F. A. Cara, G. Pedrosa, J. Meidanis and P. Arruda

- (2003). "Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane." *Genome Res* 13(12): 2725-35.
- Vicient, C. M., A. Suoniemi, K. Anamthamat-Jonsson, J. Tanskanen, A. Beharav, E. Nevo and A. H. Schulman (1999). "Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*." *Plant Cell* 11(9): 1769-1784.
- Vicient, C. M., M. J. Jaaskelainen, R. Kalendar and A. H. Schulman (2001). "Active retrotransposons are a common feature of grass genomes." *Plant Physiol* 125(3): 1283-92.
- Vicient, C. M., R. Kalendar and A. H. Schulman (2005). "Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon." *Journal Of Molecular Evolution* 61(3): 275-291.
- Vieira, C., C. Nardon, C. Arpin, D. Lepetit and C. Biemont (2002). "Evolution of genome size in *Drosophila*. is the invader's genome being invaded by transposable elements?" *Mol Biol Evol* 19(7): 1154-61.
- Vitte, C. and O. Panaud (2003). "Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L." *Mol Biol Evol* 20(4): 528-40.
- Vitte, C., C. Chaparro, H. Quesneville and O. Panaud (2007a). "Sip and Squiq, two novel rice non-autonomous LTR retro-element families related to RIRE3 and RIRE8." *Plant Science* 172(1): 8-19.
- Vitte, C., O. Panaud and H. Quesneville (2007b). "LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss." *BMC Genomics* 8: 218.
- Volff, J. N. (2006). "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes." *Bioessays* 28(9): 913-22.
- Watson, J. D. and F. H. C. Crick (1953). "Molecular structure of nucleic acids." *Nature* 171: 737-738.
- Wessler, S. R., T. E. Bureau and S. E. White (1995). "LTR-retrotransposons and MITEs: important players in the evolution of plant genomes." *Current Opinion in Genetics & Development* 5(6): 814-821.
- Wicker, T., N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf and B. Keller (2001). "Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution." *Plant J* 26(3): 307-16.
- Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel and A. H. Schulman (2007). "A unified classification system for eukaryotic transposable elements." *Nat Rev Genet* 8(12): 973-82.
- Wilhelm, M., M. Boutabout, T. Heyman and F. X. Wilhelm (1999). "Reverse transcription of the yeast *ty1* retrotransposon: the mode of first strand transfer is either intermolecular or intramolecular." *Journal of Molecular Biology* 288(4): 505-510.
- Wilhelm, M. and F. X. Wilhelm (2001). "Reverse transcription of retroviruses and LTR retrotransposons." *Cell Mol Life Sci* 58(9): 1246-62.
- Witte, C.-P., Q. H. Le, T. Bureau and A. Kumar (2001). "Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes." *PNAS* 98(24): 13778-13783.
- Wright, D. A. and D. F. Voytas (1998). "Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana* *Ty3/gypsy* retrotransposons that encode envelope-like proteins." *Genetics* 149(2): 703-715.
- Xiong, Y. and T. H. Eickbush (1990). "Origin and evolution of retroelements based upon their reverse transcriptase sequences." *Embo J* 9(10): 3353-62.

Yakushiji, H., S. Kobayashi, N. Goto-Yamamoto, S. Tae Jeong, T. Sueta, N. Mitani and A. Azuma (2006). "A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional VvmybA1 allele." *Biosci Biotechnol Biochem* 70(6): 1506-8.

Annexe

Cédric MOISY, Frédérique Pelsy*
Laboratoire Génétique et Amélioration des Plantes
UMR 1131 Santé de la Vigne et Qualité des Vins
INRA Colmar
28, rue de Herrlisheim
68021 Colmar cedex
*pelsy@colmar.inra.fr

Régulation de l'activité des éléments transposables

Les éléments transposables sont des éléments génétiques autonomes capables de se multiplier dans le génome. Ils sont en général maintenus à l'état dormant, mais certains stress peuvent induire leur transposition, c'est-à-dire leur déplacement vers un nouveau site. L'activité de ces éléments est considérée comme un facteur évolutif qui participe à l'adaptation des organismes aux variations de l'environnement.

Introduction

L'énoncé des Lois Universelles de Mendel et la possibilité d'établir des cartes génétiques se sont appuyées sur un postulat de stabilité, selon lequel les gènes sont ordonnés sur les chromosomes à des positions bien conservées au sein d'une espèce et au cours des générations. Cette théorie a ensuite été confortée par la découverte de la molécule d'ADN et de son mode de réplication semi-conservative.

Dans les années 1940, la découverte par Barbara McClintock d'éléments génétiques mobiles et autonomes chez le maïs a révélé un aspect alors encore inconnu de la génétique (1). La mobilité de ces éléments et leur capacité à se multiplier leur ont longtemps valu le qualificatif « d'ADN égoïste ». Ils ont depuis été trouvés dans tous les génomes dont ils sont parfois les constituants majeurs. A titre d'exemple, ils occupent 45% du génome humain (2) et 60% du génome du maïs (3). Les éléments transposables peuvent envahir très rapidement un génome, comme l'illustre le cas de l'élément *P*, présent dans les espèces de *Drosophila melanogaster* du monde entier alors qu'il est apparu pour la première fois il y a 40 ans (4).

Leur nombre et leur caractère invasif font inévitablement des éléments transposables des acteurs majeurs de la structure, de l'expression et de l'évolution du génome.

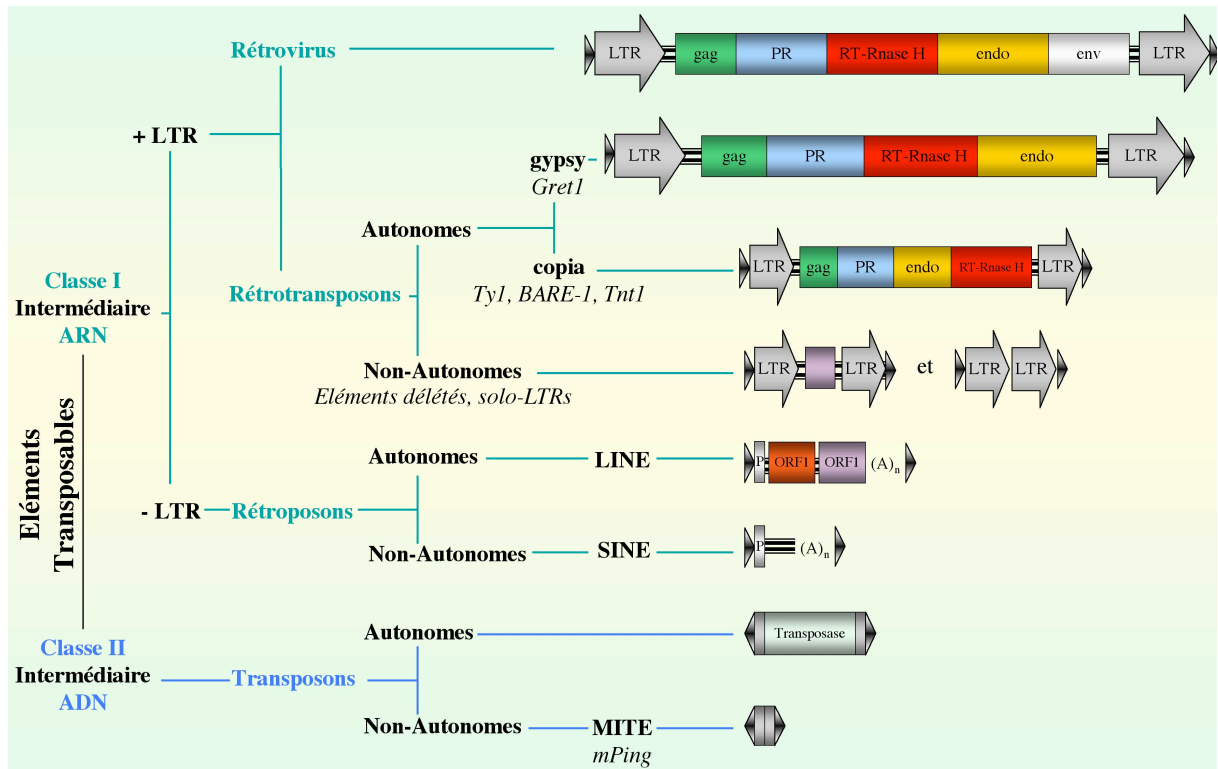


Figure 1 : Les principales classes d'éléments transposables

(A)_n : Queue Poly A+

Endo : Endonucléase

Env : gène codant pour l'enveloppe virale permettant la propagation des rétrovirus de cellule à cellule. *env* est absent chez les rétrotransposons ou subsiste à l'état non-fonctionnel

gag : code pour la nucléocapside

LINE : Long Interspersed Nuclear Elements

LTR : Long Terminal Repeat

MITE : Miniature Inverted-Repeat Transposable Elements

ORF : Open Reading Frame

P : Promoteur

▶ : séquences répétées

pol : complexe enzymatique

PR : Protéase

RT : Transcriptase Inverse

SINE : Short Interspersed Nuclear Elements

Les éléments transposables

On distingue deux classes d'éléments transposables en fonction de leur mode de transposition (figure 1).

Les éléments de la classe II, ou transposons, s'excisent de la séquence d'ADN et s'insèrent à un nouveau site, souvent proche de leur localisation initiale, par un mécanisme de type « couper-coller ». Ce cycle est contrôlé par une transposase et est souvent couplé à une amplification de la copie initiale. Les transposons sont généralement constitués de la séquence codante de la transposase, encadrée par des séquences répétées inversées.

Les éléments de la classe I, ou rétrotransposons, ont un mécanisme de transposition de type « copier-coller » qui nécessite la transcription d'une molécule d'ARNm intermédiaire puis sa transcription inverse en ADN (figure 2). Contrairement aux transposons de la classe II, la copie maîtresse du rétrotransposon reste à sa position initiale à l'issue de la rétrotransposition, seules les copies nouvellement formées s'insèrent à de nouveaux sites. Les rétrotransposons sont généralement constitués de la séquence codante des gènes « gag » et « pol », codant respectivement pour la nucléocapside et pour le complexe enzymatique. Le complexe enzymatique comprend les domaines protéase (PR), transcriptase inverse (RT), RnaseH et endonucléase (endo). Cette séquence est bordée de deux répétitions orientées dans le même sens, appelées Long Terminal Repeats (LTR), qui contiennent les promoteurs et les séquences régulatrices de la transcription (figure 1). L'ARNm transcrit à partir de l'élément maîtresse a tout d'abord un rôle dans la synthèse des protéines nécessaires au cycle de transposition. Il sert ensuite de matrice à la transcription inverse de nouvelles copies ADN de l'élément qui pourront s'intégrer au génome (figure 2).

Le cycle des rétrotransposons s'apparente à celui des rétrovirus, car il nécessite la formation dans le cytoplasme d'une particule de type viral (VLP pour Virus Like Particule) dans laquelle la transcription inverse a lieu. Mais à l'inverse des rétrovirus, les rétrotransposons ne possèdent pas de gène « env » fonctionnel, codant pour l'enveloppe nécessaire à la propagation des VLP de cellule en cellule. Le cycle de rétrotransposition n'est donc pas infectieux.

La rétrotransposition est un processus peu fidèle introduisant régulièrement des mutations dans la séquence des copies néoformées qui diffèrent alors de l'élément maîtresse d'origine. Le génome contient ainsi des « familles » d'éléments transposables, c'est-à-dire des populations de copies présentant des différences de séquence. Ces mutations, ponctuelles, insertions ou des délétions, peuvent finalement aboutir à la perte d'autonomie de l'élément. Cependant, la présence d'éléments fonctionnels dans le génome peut parfois apporter les

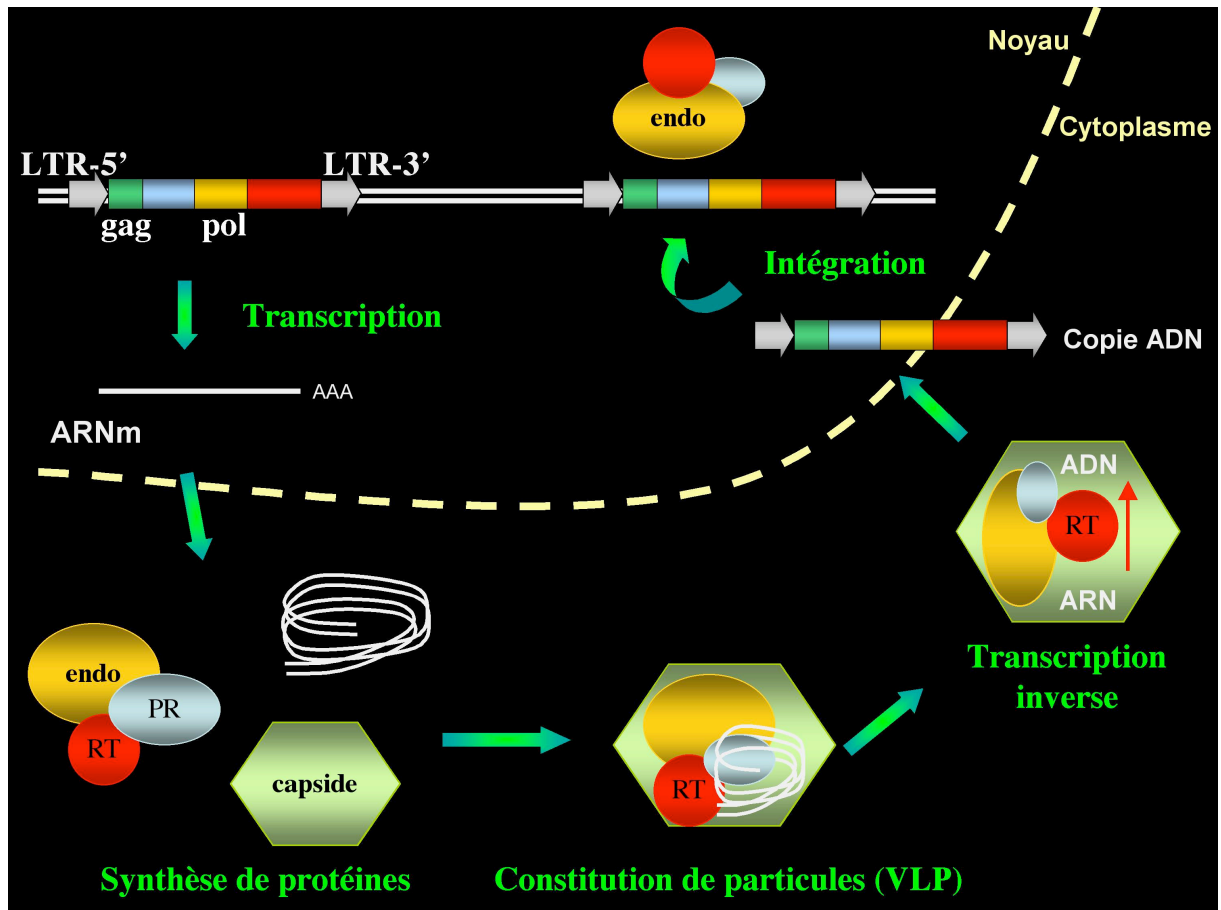


Figure 2 : Cycle de transposition d'un rétrotransposon à LTR

Le mécanisme de rétrotransposition ressemble à celui de la réplication des rétrovirus. Le gène gag code pour les protéines qui formeront la VLP (Particules de Type Viral). Le gène pol code pour la machinerie enzymatique contenant les domaines protéase (PR), transcriptase inverse (RT), endonucléase (endo). Une fois dans le cytoplasme de la cellule, l'ARN transcrit à partir du rétrotransposon est encapsidé dans la VLP. Des copies ADN y sont synthétisées grâce à l'activité transcriptase inverse. L'intégration d'une nouvelle copie dans le génome se fait ensuite au moyen de l'activité endonucléase/intégrase et d'un mécanisme de clivage/ligation. LTR : Long Terminal Repeat

protéines manquantes et permettre ainsi aux éléments mutés d'accomplir leur cycle. Ainsi, une nouvelle insertion du rétrotransposon *Zeon* qui ne possède pas la séquence codante de la transcriptase inverse a été découverte chez le maïs (5). De même, l'homologie de séquences entre les familles LINE et SINE permet de supposer que la mobilité des éléments de type SINE, qui sont délétés pour la séquence codante (figure 1), est assurée par les éléments LINE possédant une transcriptase inverse fonctionnelle.

Impact sur le génome

Les conséquences de la mobilité des éléments transposables et leur aptitude à créer des sites de recombinaison en font des acteurs importants de l'évolution des génomes. Ils sont en effet capables d'en modifier la structure, la dynamique et l'expression.

Certains éléments transposables ont un rôle structural en contribuant au maintien et à l'évolution de l'organisation des chromosomes. C'est le cas du rétrotransposon *gypsy* qui participe à l'organisation nucléaire du génome de la *Drosophile* en s'associant à des protéines pour former des isolateurs chromatinien. Ces structures stabilisent la chromatine en isolant l'hétérochromatine (ADN condensé) de l'euchromatine et ancrent les chromosomes à la membrane nucléaire en individualisant des boucles (6). A l'intérieur d'une boucle, des interactions moléculaires sont possibles, entre un promoteur et un enhancer par exemple. Un élément *gypsy* inséré à une nouvelle position peut alors isoler deux régions précédemment situées sur la même boucle, et ainsi empêcher certaines interactions moléculaires.

Chez les eucaryotes, les centromères sont constitués de séquences hautement répétées qui peuvent être associées à des rétrotransposons spécifiques, comme le rétrotransposon *CRR* dans les centromères des chromosomes du riz (7). De même, la stabilité des télomères de la *Drosophile* est assurée par la répétition de motifs constitués par les deux rétrotransposons, *HeT-A* et *TART*, associés en tandem (8).

Succédant aux phases d'amplification, l'accumulation de délétions par des recombinaisons impliquant deux copies d'un même élément peut conduire à l'élimination rapide de ces séquences, ou à la formation de copies filles fortement délétées. L'insertion de copies homologues de part et d'autre d'un gène ou d'un exon peut en effet aboutir à des remaniements chromosomiques et à l'excision de toute l'unité, générant un nouvel arrangement de la structure génique.

Bien que l'insertion d'une nouvelle copie d'élément transposable soit en règle générale sans conséquence pour le génome hôte, elle reste une cause importante de mutation. Lorsqu'elle a

lieu dans la séquence codante, l'insertion peut bloquer l'expression du gène ou en modifier la fonction. A titre d'exemple, les éléments transposables sont impliqués dans 50% des mutations délétères chez la *Drosophile* (9).

Si l'élément s'insère dans la séquence régulatrice du gène, ou s'il éloigne le promoteur de la séquence codante qu'il contrôle, l'expression du gène peut en être perturbée. Chez la vigne, l'insertion du rétrotransposon *Gret1* en amont du gène *MybA* bloque la synthèse des anthocyanes dans les variétés à raisin blanc (10).

Cependant, l'insertion d'un élément transposable n'est pas systématiquement délétère. La présence de promoteurs et de terminateurs dans les LTR des rétrotransposons peut également influencer la régulation des gènes localisés à proximité, voire se substituer aux promoteurs spécifiques présents initialement. Dans le génome humain, 25% des gènes sont localisés à proximité d'un élément transposable susceptible d'en moduler l'expression, et pour 3% des gènes humains annotés, les rétrotransposons apportent les signaux de régulation de la transcription (11).

L'utilisation de promoteurs dérivés d'éléments transposables peut parfois avoir un rôle dans l'expression tissu-spécifique de certains gènes. De plus, une insertion antisens par rapport à un gène peut permettre la synthèse d'un transcrit pouvant réprimer l'expression du gène par un phénomène d'interférence des ARN.

Régulation de la transposition

Les effets délétères résultant de la mobilité des éléments transposables présentent un risque pour la survie de l'hôte et donc pour le maintien des éléments eux-mêmes. Cependant, leur ubiquité suggère qu'ils ont développé une stratégie leur permettant de coexister dans le génome hôte, voire de lui être bénéfique. L'hôte limite le potentiel délétère des éléments transposables en contrôlant leur activité par différents mécanismes.

Chez le riz, il a été montré qu'en parallèle de l'amplification de quelques copies maître, les séquences nouvellement insérées ont été éliminées plus ou moins rapidement par recombinaison. De plus, les mutations introduites dans la séquence des éléments transposables provoquent souvent la perte de leur autonomie. Par ailleurs, des mécanismes épigénétiques tels que la méthylation et la condensation de la chromatine peuvent également participer à la régulation de leur activité.

La transposition des éléments mobiles peut potentiellement être régulée à chaque étape du cycle. Au niveau transcriptionnel, la méthylation des éléments transposables est corrélée à

leur inactivation chez le maïs (12). Dans le cytoplasme, la dégradation spécifique des transcrits par interférence ARN est un autre mécanisme qui peut inhiber le cycle de transposition des rétrotransposons.

Chez la levure, les transcrits du rétrotransposon *Ty1* représentent près de 0,8% des ARN totaux de la cellule, mais la maturation des protéines de *Ty1* semble très inefficace et les VLP ne sont présentes qu'en très faible quantité (13). Finalement, une seule copie d'ADN est synthétisée en moyenne pour 14000 transcrits *Ty1* car des protéines de l'hôte inhibent les étapes post-transcriptionnelles de la rétrotransposition (14).

Le contrôle transcriptionnel des éléments transposables se traduit par leur faible représentation dans les collections de séquences exprimées (EST expressed sequence tags). Dans une collection de 407 000 EST de maïs, seuls 56 dérivent de rétrotransposons à LTR alors que ces éléments sont des constituants majeurs du génome (15).

Via l'activité de ses éléments transposables, le génome hôte bénéficie d'un « réservoir mutagène » qui lui confère un potentiel adaptatif pour faire face aux changements environnementaux. De plus, les éléments transposables apportent à leur hôte un répertoire protéique présentant des propriétés fonctionnelles intéressantes d'excision, de réplication et d'intégration de l'ADN dont il peut profiter s'il parvient à les domestiquer.

Les stress activent la mobilité des éléments transposables

Qu'ils soient ponctuels ou permanents, les stress biotiques (attaque de virus, de champignons...) ou abiotiques (blessure, variations de température, sécheresse...) exercent une pression à laquelle les organismes doivent s'adapter pour survivre. L'adaptation des mécanismes physiologiques, la régulation des gènes, ainsi que les remodelages génomiques sont autant de mécanismes mis en œuvre en réponse aux stress (16). Du point de vue évolutif, l'activation des éléments transposables apporte une « plasticité génétique » qui favorise l'évolution de l'organisme soumis au stress (17). Différentes études ont montré que les stress augmentent la variabilité génétique, en particulier par l'accroissement du nombre de mutations d'insertion impliquant les éléments transposables. L'environnement et les stress affectant l'hôte ont donc une répercussion sur le niveau de transposition.

A titre d'exemple, une relation entre le nombre de copies de *BARE-1* et le microclimat a été établie en étudiant les individus des versants Nord et Sud d'un canyon en Israël (18). Le polymorphisme d'insertion permet de distinguer très clairement les deux populations pourtant

distantes de 300 mètres au maximum. Le nombre de copies de *BARE-1* est plus grand dans les plantes poussant en altitude et en condition aride, particulièrement au sommet du versant Sud du canyon. La présence, dans la séquence promotrice des LTR de *BARE-1*, d'un motif de réponse à l'acide abscissique, caractéristique des gènes induits par le stress hydrique, pourrait expliquer la corrélation entre l'induction de *BARE-1* et les conditions de sécheresse.

Chez le riz, l'amplification du nombre de copies du transposon *mPing* est plus importante dans les variétés adaptées aux climats tempérés (19). Ces variétés ont été sélectionnées pour leur productivité dans des conditions de température plus froide et de saison de culture plus courte. Il est possible que l'activation des *mPing* par le stress subi lors du changement de climat, suivie de leur insertion dans les régions codantes, ait participé à la diversification de ces variétés.

A la suite de son insertion dans le gène de la nitrate réductase (NR) du tabac, l'élément *Tnt1* a été le premier rétrotransposon caractérisé chez les plantes (20). Le protocole de sélection des mutants déficients pour la NR nécessitait la préparation de protoplastes et donc l'utilisation d'extraits fongiques pour dégrader les parois cellulaires. Il a été démontré par la suite que les éliciteurs fongiques libérés lors de la lyse des parois étaient en fait à l'origine de l'activation de *Tnt1* lors de la préparation des protoplastes, et de son insertion dans le gène de la NR.

L'expression de *Tnt1* est étroitement régulée en conditions normales de croissance de la plante et n'est pas détectée dans la plupart des tissus, à l'exception des racines adultes (21). Trois sous-familles de *Tnt1* ont été décrites, toutes transcrites en réponse à des stress (22). La transcription de *Tnt1A* est fortement induite dans les feuilles traitées avec des éliciteurs fongiques (cryptogéine, méthyl jasmonate). *Tnt1B* est induit dans les cultures cellulaires et *Tnt1C* par le traitement à l'acide salicylique.

Ces résultats suggèrent un lien entre la réponse de la plante au stress et l'activation de la transcription de *Tnt1* via l'activation de gènes impliqués dans les mécanismes de défense (20). Les promoteurs de *Tnt1* présentent des homologies avec certains promoteurs de gènes de défense et sont reconnus par les mêmes facteurs de transcription (22). De la même manière, un motif de 13 paires de bases, interagissant avec plusieurs facteurs de transcription induits par infection virale, a été retrouvé dans le promoteur des LTRs de *Tto1*, un autre rétrotransposon du tabac induit par des blessures.

Les promoteurs de *Tnt1* n'ayant pas évolué indépendamment du reste de la séquence du rétrotransposon, les similarités observées avec des gènes de défense semblent résulter d'une évolution convergente (23).

Conclusion

En montrant leur implication dans l'expression, la structure et la dynamique des génomes, les résultats de l'étude des éléments transposables ouvrent des perspectives pour la compréhension des mécanismes qui gouvernent l'évolution des génomes. En conditions de stress, la réponse de l'organisme n'est pas strictement transcriptionnelle mais aussi structurale grâce à la mobilité des éléments transposables qui assurent une certaine plasticité aux génomes.

Au Laboratoire de Génétique et d'Amélioration des Plantes de l'INRA de Colmar, l'impact de l'activité des rétrotransposons sur la genèse de la diversité intravariétale chez la vigne est évalué. Les variétés de vigne sont composées d'un ensemble de clones issus de la même plante propagée par multiplication végétative sur de longues périodes. Bien que le processus de multiplication permette le plus souvent d'assurer le maintien de l'identité génétique et phénotypique des clones, des mutations spontanées peuvent apparaître et être à l'origine de nouveaux types de clones présentant des aptitudes culturales et technologiques différentes de la plante initiale. De cette manière, la famille des pinots s'est diversifiée en pinot noir, pinot blanc et pinot gris, chacun permettant la production d'un vin bien caractéristique (24).

La vigne est une plante pérenne soumise à de nombreux stress au cours de sa période de culture : sécheresse, exposition aux rayonnements solaires, blessures répétées liées aux conditions de culture (greffage, taille, rognage) et aux attaques de nombreux pathogènes. Nous cherchons à mettre en évidence dans quelle mesure les éléments transposables participent à l'évolution du génome des variétés de vigne à l'origine de la diversité intravariétale.

1. McClintock, B. (1984) *Science*, 226, 792-801
2. Lander, E.S.e.a. (2001) *Nature*, 409, 860-921
3. SanMiguel, P., et al. (1996) *Science*, 274, 765-768
4. Anxolabehere, D., et al. (1988) *Mol Biol Evol*, 5, 252-269
5. Hu, W., et al. (1995) *MGG*, 2484, 471-480
6. Byrd, K. and V.G. Corces (2003) *The Journal Of Cell Biology*, 1624, 565-574
7. Cheng, Z., et al. (2002) *The Plant Cell*, 148, 1691-1704
8. Pardue, M.-L. and P.G. DeBaryshe (2003) *Annual Review Of Genetics*, 37, 485-511
9. Finnegan, D.J. (1992) *Current Biology*, 212, 632
10. Kobayashi, S., et al. (2004) *Science*, 304, 982
11. Van de Lagemaat, L.N., et al. (2003) *TIG*, 1910, 530-536
12. Anxolabehere, D., et al. (2000) *Société Francaise et génétique*, 1611, 1-9
13. Curcio, M.J., et al. (1990) *MGG*, 2202, 213-221
14. Curcio, M.J. and D.J. Garfinkel (1999) *TIG*, 152, 43-45
15. Meyers, B.C., et al. (2001) *Genome Research*, 1110, 1660-1676
16. Madlung, A. and L. Comai (2004) *Ann Bot*, 944, 481-495
17. Melayah, D., et al. (2001) *The Plant Journal*, 282, 159-168
18. Kalendar, R., et al. (2000) *PNAS*, 97, 6603-6607
19. Jiang, N., et al. (2003) *Nature*, 421, 163-167
20. Grandbastien, M.-A., et al. (1997) *Genetica*, 1-3, 241-252
21. Pourtau, N., et al. (2003) *Plant Science*, 1655, 983-992
22. Casacuberta, J.M. and N. Santiago (2003) *Gene*, 311, 1-11
23. Vernhettes, S., et al. (1998) *Mol Biol Evol*, 157, 827-836
24. Hocquigny, S., et al. (2004) *Genome*, 473, 579-589

Abstract

Grapevine (*Vitis vinifera* L.) varieties are composed by clones originated from a unique plant propagated by vegetative multiplication. This process theoretically ensure the transmission of genetic information from mother to daughter's plants, but spontaneous mutations sometimes arise and can lead to the creation of new genotypes. Due to their ability to transpose, transposable elements are a source of mutation that causes genetic variability. To understand global mechanisms that participate to plant genome evolution during vegetative multiplication cycles, the main goal of this thesis is studying the structure and dynamics of grapevine retrotransposons.

When this work started, only three grapevine retrotransposons were described. Therefore, the first part of this work is based on the characterization of ten new families, which are different from the elements previously described in grapevine or in other plant genomes. This ten new families represent near 1,700 copies, from full-length to degenerated, representing altogether about 1.2% of the genome. Some of these copies were integrated in the grapevine genome during the last two million years. In a second part, the transcription level of grapevine retrotransposons is evaluated by complementary approaches. The results show that their transcription level is regulated in different organs under normal growing conditions, each family showing a specific transcription profile. Moreover, the expression of many retrotransposons in leaves is induced by wounding and pathogen infection.

The role of retrotransposons in grapevine genome evolution and in the genesis of intra-varietal diversity has not been established, but induction of expression of several families under stress conditions allow us to suppose that this elements might have participated to genome adaptation in response to environmental challenges and grapevine culture by man. Moreover, the characterization of new transposable families contributes to the development of molecular tools allowing the identification and distinction of grapevine clones.

Key words: *Vitis vinifera* L., genetic diversity, retrotransposon, mobile elements, stress, transcription, RT-PCR

Résumé

Les variétés de vigne (*Vitis vinifera* L.) sont composées d'un ensemble de clones, tous issus de la même et unique plante propagée par multiplication végétative sur de longues périodes. Ce processus doit en principe assurer la transmission du patrimoine génétique de la plante mère à la plante fille, cependant des mutations spontanées peuvent apparaître et être, dans certains cas, à l'origine de nouveaux génotypes. Les éléments transposables, par la variabilité génétique que peut générer leur mobilité, sont une source potentielle de mutation. Dans un contexte général de compréhension des mécanismes biologiques qui participent à l'évolution des génomes au cours des cycles de multiplication végétative, l'objectif de la thèse est d'étudier la structure et la dynamique des rétrotransposons du génome de la vigne.

Au début de la thèse, trois rétrotransposons du génome de la vigne étaient décrits. Une première partie de ce travail est consacrée à la caractérisation de dix nouvelles familles de rétrotransposons qui se distinguent des trois éléments déjà connus chez la vigne, ainsi que des rétrotransposons décrits chez d'autres plantes. L'ensemble de ces familles comptent environ 1700 copies, complètes et potentiellement actives ou partiellement dégénérées, dont certaines se sont insérées dans le génome de la vigne au cours des deux derniers millions d'années. Au total, ces familles représentent près de 1,2 % du génome. Dans une deuxième partie, l'étude du niveau de transcription de ces rétrotransposons par des approches complémentaires montre que la présence de leurs transcrits est régulée dans les organes de la vigne en conditions normales de croissance, chaque famille d'éléments présentant un profil d'expression particulier selon l'organe considéré. De plus, l'expression de plusieurs rétrotransposons est induite dans les feuilles en réponse à la blessure et à l'infection par le mildiou.

Le rôle des rétrotransposons dans l'évolution du génome de la vigne et la genèse de diversité intra-variétale reste encore à déterminer. Cependant, l'induction de l'expression de différentes familles en condition de stress permet d'envisager que ces éléments pourraient avoir participé à l'adaptation du génome en réponse aux contraintes culturales et aux conditions environnementales. De plus, la caractérisation de nouvelles familles d'éléments, dont certaines sont actives et présentes en un grand nombre de copies, peut permettre le développement d'outils moléculaires adaptés à l'identification et à la distinction des clones de vigne.

Mots-clés : *Vitis vinifera* L., diversité intra-variétale, rétrotransposon, éléments mobiles, stress, transcription, RT-PCR