



**HAL**  
open science

# Expression et évolution de gènes de la famille des ARN hélicases à boîte DEAD chez *Arabidopsis thaliana* (L.)

Heynh

Annaïck Mingam

► **To cite this version:**

Annaïck Mingam. Expression et évolution de gènes de la famille des ARN hélicases à boîte DEAD chez *Arabidopsis thaliana* (L.) Heynh. Sciences du Vivant [q-bio]. Université Paris Sud - Paris XI, 2004. Français. NNT: . tel-00007807

**HAL Id: tel-00007807**

**<https://theses.hal.science/tel-00007807>**

Submitted on 17 Dec 2004

**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.

ORSAY  
N° d'ordre : 7530

UNIVERSITE DE PARIS-SUD  
U.F.R. SCIENTIFIQUE D'ORSAY

THESE

**présentée  
pour obtenir le grade de**

DOCTEUR EN SCIENCES  
DE L'UNIVERSITE PARIS 11, ORSAY

*Discipline : Biologie Végétale*

par

**Annaïck MINGAM**

Soutenance : le 8 juin 2004

**Expression et évolution de gènes de la famille  
des ARN hélicases à boîte DEAD  
chez *Arabidopsis thaliana* (L.) Heynh**

Directeur de thèse : Dr Alain LECHARNY

Jury :

Françoise Vedele	Rapporteur
Richard Cooke	Rapporteur
Michel Termier	Examineur
Pierre Gadal	Examineur
Alain Lecharny	Directeur de thèse

## Remerciements

Je remercie Martin Kreis, directeur de l'Institut de Biotechnologie des Plantes et directeur du Laboratoire « Génome, évolution et développement de la plante », de m'avoir accueillie dans son équipe et soutenue pendant toute la durée de ma thèse.

Je remercie Alain Lecharny d'avoir encadré ma thèse avec rigueur scientifique, persévérance et ténacité.

Merci à Michel Dron, directeur de l'Ecole doctorale, et à Pierre Gadal d'avoir suivi avec attention le déroulement de ma thèse.

J'adresse un grand merci à Jean-Paul Barès et Gilles Santé de s'être occupés de mes plantes, à Isabelle Gy et Gaëlle Claisse pour la qualité du service séquençage et à Roland Boyer pour son travail de photographe et sa disponibilité.

Les remerciements sont un exercice de style dans lequel il faut à la fois respecter un certain nombre de contraintes tout en y mettant sincérité et émotion. Pour moi, l'objectif de ce paragraphe va au-delà des simples remerciements en tant que tels et permet à son auteur comme aux lecteurs de mettre des mots sur la fin d'un travail et d'un moment partagé. Alors, pour que la page se tourne dans la joie et que la route continue belle et ensoleillée pour tout le monde, j'adresse mes remerciements chaleureux, sautillants, vivants et enthousiastes à Alain, Martin, Claire, Nicole, Cécile, Nicolas, Gaëlle, Yves, Bénédicte, Odile, Mathilde, Soazig, Céline, Dao, Nicolas, Martine, Mathieu, Isabelle, Claire, Moussa, Nathalie, Lionel, Anthony, Marianne, Thomas, Pierre, Arnaud, Darko, Hélène, Alain, Claire, Micheline, Sébastien, Laure, Clémence, Mariette et Vincent, Ting, mes amis, mes parents, mes frères et ma sœur, ma grand-mère pour toute l'énergie déployée qui a rendu cette thèse possible.

### **Remarques générales**

Dans ce manuscrit, pour éviter de lourdes périphrases, j'ai été amenée à utiliser occasionnellement une acception non orthodoxe du mot « gène » à la place des mots transcrits, ARNm ou ADNc. De même, l'expression ou l'expression transcriptionnelle d'un gène font référence à la teneur des transcrits de ce gène dans un échantillon donné, valeur qui résulte à la fois de l'activité du promoteur et de la vitesse de dégradation de ces transcrits.

Par ailleurs, les mots étrangers, provenant notamment du latin et de l'anglais, sont écrits en italiques. Les noms des protéines sont écrits en majuscules et les noms des gènes en majuscules et en italiques.

## Abréviations

<b>ABA</b>	Acide abscissique	
<b>ADN</b>	Acide désoxyribonucléique	
<b>ADNc</b>	ADN complémentaire	
<b>AFLP</b>	<i>Amplified Fragment Length Polymorphism</i>	Polymorphisme de longueur des fragments amplifiés
<b>AGI</b>	<i>Arabidopsis Genome Initiative</i>	Initiative pour le génome d' <i>Arabidopsis</i>
<b>ARN</b>	Acide ribonucléique	
<b>ARNm</b>	ARN messenger	
<b>ARNr</b>	ARN ribosomique	
<b>AtRH</b>	ARN hélicase d' <i>Arabidopsis thaliana</i>	
<b>BLAST</b>	<i>Basic Local Alignment Search Tool</i>	Outil de recherche d'alignements locaux
<b>BrEt</b>	Bromure d'éthidium	
<b>CeSC</b>	<i>Caenorhabditis elegans Sequencing Consortium</i>	Consortium pour le séquençage de <i>C. elegans</i>
<b>Ct</b>	<i>Threshold cycle</i>	Cycle-seuil
<b>dCTP</b>	Désoxycytosine triphosphate	
<b>dUTP</b>	Désoxyuracile triphosphate	
<b>dTTP</b>	Désoxythymidine triphosphate	
<b>EDTA</b>	Ethylènediamine tétraacétique	
<b>EST</b>	<i>Expressed Sequence Tag</i>	Etiquette de séquence transcrite
<b>eg</b>	Equivalent-génome	
<b>FRET</b>	<i>Fluorescence Resonance Energy Transfer</i>	Transfert d'énergie de la fluorescence par résonance
<b>GST</b>	<i>Gene Sequence Tag</i>	Etiquette de séquence génique
<b>IAA</b>	<i>Indole-3-acetic acid</i>	acide indole-3-acétique (auxine)
<b>MA</b>	Million d'années	
<b>MPSS</b>	<i>Massively Parallel Signature Sequencing</i>	Séquençage massif de signatures en parallèle
<b>nt</b>	Nucléotide	
<b>OGM</b>	Organisme génétiquement modifié	
<b>ORF</b>	<i>Open Reading Frame</i>	Cadre ouvert de lecture
<b>pb</b>	Paire de bases	
<b>PCR</b>	<i>Polymerase Chain Reaction</i>	Réaction de polymérisation en chaîne
<b>QTL</b>	<i>Quantitative Trait Locus</i>	Locus de trait quantitatif
<b>RSAT</b>	<i>Regulatory Sequences Analysis Tool</i>	Outil d'analyse de séquences régulatrices
<b>RT</b>	<i>Reverse Transcription</i>	Transcription inverse
<b>SAGE</b>	<i>Serial Analysis of Gene Expression</i>	Analyse en série de l'expression des gènes
<b>ScrH</b>	ARN hélicase de <i>Saccharomyces cerevisiae</i>	
<b>SDS</b>	<i>Sodium dodecyl sulfate</i>	
<b>UTR</b>	<i>Untranslated region</i>	Région non codante
<b>UV</b>	Ultraviolet	
<b>Tm</b>	Température de fusion	
<b>Tris</b>	Tris hydroxyméthylaminométhane	
<b>UNG</b>	Uracyl-N-glycosylase	

## Plan

<b>Remerciements</b>	<b>1</b>
<b>Remarques générales</b>	<b>2</b>
<b>Abréviations</b>	<b>3</b>
<b>Plan</b>	<b>4</b>
<b>Introduction</b>	<b>9</b>
<i>I. Le génome d'Arabidopsis thaliana et sa dynamique</i>	<i>9</i>
A. <i>Arabidopsis thaliana</i> : plante et génome modèle	9
1. Plante modèle	9
2. Génome modèle	9
3. Un génome dupliqué	10
B. Origine des duplications	10
1. Les duplications de grands fragments	10
2. Les autres duplications	11
C. L'émergence des familles multigéniques	12
1. La formation d'une famille	12
2. Les mécanismes d'évolution d'une famille	12
a. La sélection positive	12
b. Le modèle duplication-divergence-complémentation	13
3. Conséquences fonctionnelles : redondance ou diversité ?	13
<i>II. Problématique</i>	<i>15</i>
A. La régulation de l'expression comme moteur de l'évolution fonctionnelle	15
B. Intérêts de l'étude de l'expression transcriptionnelle d'une famille multigénique	15
C. Objectifs	17
<i>III. Les ARN hélicases à boîte DEAD : un modèle d'étude</i>	<i>18</i>
A. Présentation de la famille	18
B. Modèle d'étude de l'évolution de la régulation de l'expression transcriptionnelle	22
<i>IV. De l'étude du transcriptome à l'étude de l'expression transcriptionnelle d'une famille</i>	<i>23</i>
A. Les moyens d'étude de l'expression transcriptionnelle à grande échelle	23
1. L'étiquetage des transcrits	23
a. Les EST, étiquettes de séquences exprimées	23
b. Les étiquettes SAGE	25
c. La technique MPSS	26
d. Conclusions	26
2. ADNc-AFLP	28
3. Les puces à ADN	28
a. Les puces d'ADNc ou de produits PCR	28
b. Les puces d'oligonucléotides	30
c. Les puces représentant le génome entier	30
1. Les méthodes d'analyse globale : étiquettes et AFLP	33

Plan	5
2. Les puces à ADN	33
3. La PCR quantitative en temps réel	34
<b>Résultats</b>	<b>36</b>
<b>La PCR quantitative en temps réel</b>	<b>37</b>
<i>I. Introduction</i>	37
<i>II. Utilisations de la PCR quantitative en temps réel en biologie végétale</i>	42
A. Sécurité alimentaire : détection d'OGM et de contaminants fongiques et bactériens	42
B. Biologie Moléculaire	43
1. Gènes et transcrits végétaux utilisés comme standards	43
2. Transgenèse	43
3. Outil de validation	43
C. Phytopathologie	44
<i>III. Principe et définitions</i>	45
A. Dispositif expérimental	45
B. Quantification de l'ADN par la détection de fluorescence	47
1. Cycle-seuil et nombre de copies initial	47
2. Equation de la gamme-étalon	48
3. Relation pente-efficacité de la PCR	48
4. Efficacité des amorces et variations expérimentales de la gamme étalon	49
C. Choix expérimentaux	50
1. Le système d'émission de fluorescence	50
2. Gamme-étalon	50
3. Interprétation des résultats en nombre de copies	53
4. Choix des amorces	53
<i>IV. Validation de la PCR quantitative en temps réel dans le cadre de l'étude d'une famille de gènes</i>	57
A. Etendue de la gamme dynamique	57
B. Spécificité des amorces	59
C. Stratégie expérimentale	60
D. Reproductibilité technique et biologique	61
E. Qualité des échantillons	64
<i>V. Conclusion</i>	66
<b>Expression des <i>AtRH</i> et analyse des promoteurs</b>	<b>67</b>
<i>I. Introduction</i>	68
A. Choix des gènes	71
B. Niveaux d'expression transcriptionnelle	73
1. PCR quantitative en temps réel	73
2. <i>Northern blots</i>	77
a. Expression et nombre d'EST	81
b. Expression et MPSS	85
C. Profils d'expression transcriptionnel	87
D. Conclusion	89
<i>III. Eléments cis-régulateurs et niveaux de transcription</i>	91

A. Classification des <i>AtRH</i> en fonction de la structure des régions promotrices	91
1. Annotation structurale des <i>AtRH</i>	91
2. Identification de la boîte TATA	91
B. Groupes structuraux et niveaux d'expression des gènes <i>AtRH</i>	93
IV. <i>Etude des promoteurs</i>	96
A. Caractérisation des séquences promotrices	96
B. Représentation en 3 dimensions	96
C. Analyse des promoteurs et de la région 5' UTR avec le logiciel RSAT	96
1. Deux motifs sur-représentés dans le promoteur des <i>AtRH</i>	97
2. Analyse à l'échelle du génome	97
a. Abondance des boîtes télo et GGCCCA	97
b. Localisation et répartition des boîtes dans les promoteurs	101
3. Lien entre expression et boîtes	103
V. <i>Groupes d'orthologues et transcription</i>	105
<b>Discussion</b>	<b>107</b>
I. <i>Quantification de la teneur des ARNm par RT-PCR quantitative en temps réel</i>	107
A. Recommandations	107
B. La normalisation des résultats	108
C. Sensibilité et spécificité	109
1. Sensibilité	109
2. Spécificité	110
D. Comparaison de la RT-PCR quantitative au <i>Northern blot</i>	111
1. Sensibilité	111
2. Spécificité	112
3. Limites et complémentarités	112
E. Une technique complémentaire aux <i>microarrays</i>	114
II. <i>Expression de familles de gènes</i>	116
A. Familles de gènes « de maintenance » et expression constitutive	116
B. Divergence des profils	117
C. Divergence des niveaux d'expression	118
1. Proximité des séquences protéiques et niveau d'expression	118
2. Niveau d'expression et rôle biologique	119
III. <i>Analyse bioinformatique des promoteurs</i>	121
A. Fonctions des boîtes télo et GGCCCA	121
1. Définitions	121
2. Les boîtes télo et GGCCCA	121
B. D'autres éléments dans les introns	123
<b>Perspectives</b>	<b>125</b>
I. <i>Apport de l'approche « analyse d'une famille » à la connaissance du génome</i>	125
II. <i>Importance de la topologie dans l'analyse des promoteurs</i>	126
III. <i>Néo-fonctionnalisation et sous-fonctionnalisation</i>	127
<b>Matériel et Méthodes</b>	<b>128</b>



<i>I. Biologie moléculaire</i>	128
A. Matériel biologique	128
1. Culture	128
2. Prélèvements	128
B. Manipulation d'acides nucléiques	128
1. Extraction	128
a. ADN génomique d' <i>A. thaliana</i>	128
b. ARN d' <i>A. thaliana</i>	129
c. ADN plasmidique	129
2. Précipitation de l'ADN	130
3. Dosage et stockage des acides nucléiques	130
C. Synthèse d'ADNc simple brin	130
D. Amplification par PCR	131
1. PCR classique	131
2. PCR quantitative en temps réel	132
a. Conditions expérimentales	132
b. Cycles thermiques	132
E. Analyse des produits d'amplification : électrophorèse en gel d'agarose	132
F. <i>Dot Blot</i>	133
1. Principe	133
2. Hybridation	133
3. Purification des sondes radiomarquées	134
4. Préparation de la résine Sephadex G-50 pour la purification des sondes radioactives par chromatographie par exclusion	134
G. Séquences	134
<i>II. Bioinformatique</i>	136
A. Analyse de séquences	136
1. Recherche de similarités	136
2. Alignement de séquences	136
a. Alignement multiple	136
b. Alignement de transcrits	137
3. Outil de visualisation	137
B. Outils informatiques pour le choix des couples d'amorces	137
1. Oligo 4.0	138
2. Primer Express™	138
3. Amplify 1.2	138
C. Traitements des données de PCR quantitative	139
1. GeneAmp 5700SDS	139
2. Macros Excel	139
D. Analyse des promoteurs	141
1. Séquences	141
2. Analyse de courbure	141
3. Recherche de motifs dans les promoteurs	142
a. PlantCARE	142
b. RSAT, outil d'analyse de séquences régulatrices	142
<b>Bibliographie</b>	<b>144</b>
<b>Annexes</b>	<b>162</b>

<i>Annexe 1</i>		<b>163</b>
	<b>Mingam A, Gachon C and Charrier B.</b> Quantifying nucleic acids with fluorescent probes: contribution of real-time PCR to applied and fundamental plant research. Review. <i>Advances in Plant Physiology</i> . 7 : 423-458.	
<i>Annexe 2</i>		<b>198</b>
	<b>Gachon C, Mingam A and Charrier B.</b> Real-time PCR: what relevance to plant studies? Review. <i>Journal of Experimental Botany</i> . Sous presse.	
<i>Annexe 3</i>	Code de la macro « Modifie1_TraitementFichierCSV »	<b>209</b>

## Introduction

### I. Le génome d'*Arabidopsis thaliana* et sa dynamique

#### A. *Arabidopsis thaliana* : plante et génome modèle

##### 1. Plante modèle

Comme son nom vernaculaire, champêtre ou empreint de romantisme, ne l'indique pas, l'arabette des dames ou des prés est une petite crucifère au centre des intérêts scientifiques. Bien que poussant à l'état sauvage dans nos régions tempérées, je l'ai, pour ma part, majoritairement croisée dans les serres et les salles de cultures des instituts de recherche en biologie végétale. Comment expliquer le succès de cette « mauvaise herbe » auprès des chercheurs ? Ni ses qualités nutritives, ni ses qualités ornementales sont à mettre en cause. Au contraire, ce sont ses caractéristiques typiques de « mauvaise herbe » qui lui ont donné son charme.

Tout d'abord, *Arabidopsis thaliana* (L.) Heynh présente une croissance rapide et une grande fertilité. En conditions de culture optimales, deux mois suffisent pour réaliser un cycle de reproduction complet et obtenir plusieurs milliers de graines à partir d'une seule. Sa petite taille (inférieure à 40 cm) permet de cultiver une centaine de plants par m<sup>2</sup> sans altérer sa croissance. Bien que l'autofécondation soit le mode naturel de reproduction de cette plante, l'allofécondation forcée est possible. Ces nombreux avantages ont fait de cette plante un outil privilégié des généticiens, dès le début des années soixante, pour caractériser des mutants affectés dans des voies métaboliques, dans le développement ou la reproduction (Estelle & Somerville, 1986). De nombreux mutants ont été générés et étudiés et ont ainsi contribué à la connaissance actuelle de la physiologie d'*A. thaliana*.

##### 2. Génome modèle

Dans les années 90, l'épopée génomique commençait, les premiers programmes de séquençage de génomes complets étaient lancés et il fallait choisir un modèle pour le règne végétal. Une nouvelle fois *A. thaliana* fut sélectionnée. Non seulement sa culture, sa croissance, son cycle de reproduction et sa physiologie étaient bien connus, mais surtout cette dicotylédone possède l'un des plus petits génomes nucléaires des angiospermes, estimé à 120 Mb répartis en 5 chromosomes. Les données de génétique indiquaient d'autres avantages comme une faible proportion de séquences répétées (Pruitt & Meyerowitz, 1986). De plus, *A. thaliana* était considérée comme une plante diploïde (Meyerowitz, 1992) avec seulement

17% de loci dupliqués (McGrath *et al.*, 1993). Enfin, une analyse de petits fragments génomiques montrait que les séquences intergéniques étaient plutôt courtes : moins d'un kilobase environ (Le Guen *et al.*, 1994). Ainsi, le génome d'*A. thaliana* constitua le premier génome modèle des végétaux (Meinke *et al.*, 1998). Son séquençage commença en 1995 avec les programmes européens ESSA 1 et 2 pour s'achever en décembre 2000 (Goodman *et al.*, 1995; AGI, 2000).

Depuis, le riz (*Oryza sativa* L.) a été choisi comme génome modèle des monocotylédones. Le séquençage de son génome est sur le point d'être terminé (Delseny, 2003; Schoof & Karlowski, 2003).

### 3. Un génome dupliqué

Des études comparatives entre *A. thaliana* et *B. oleracea* avaient révélé certaines régions dupliquées du génome des Brassicacées (Kowalski *et al.*, 1994; Lan *et al.*, 2000). L'analyse du génome entier d'*A. thaliana* a permis de montrer que 65% des gènes sont dupliqués, avec un plus grand nombre de gènes dupliqués en tandem et par duplications segmentaires que dans le génome de *Caenorhabditis elegans* et *Drosophila melanogaster*, dont le séquençage s'est achevé en 1998 et 2000 (CeSC, 1998; Adams *et al.*, 2000; AGI, 2000). De plus, le génome d'*A. thaliana* possède un plus grand nombre de familles multigéniques que les génomes de *C. elegans* ou *D. melanogaster*. La proportion de gènes appartenant à des familles de plus de 3 membres est de 52,4% chez *A. thaliana* alors qu'elle n'est que de 19% et 32,8% dans les génomes de *D. melanogaster* et *C. elegans*, respectivement (AGI, 2000). Parallèlement, la proportion de gènes isolés n'est que de 35% dans le génome d'*A. thaliana*, contre 72,5% et 55,2% dans les génomes de *D. melanogaster* et *C. elegans*, respectivement (AGI, 2000). A une échelle plus petite, il a été constaté par exemple chez *A. thaliana* une taille plus importante de toutes les familles impliquées dans le métabolisme des ARN par rapport à celle des autres eucaryotes (Anantharaman *et al.*, 2002).

#### B. Origine des duplications

Les 65% de gènes dupliqués proviennent de différents types de duplication qui peuvent se produire au sein d'un génome : une partie de gène, un gène, un bloc de plusieurs gènes, un chromosome entier, l'ensemble du génome (Sankoff, 2001).

##### 1. Les duplications de grands fragments

Les duplications à grande échelle correspondent à des duplications plus ou moins partielles du génome. Elles peuvent concerner le génome entier (polyploïdisation), des chromosomes (aneuploïdisation) ou des fragments de chromosome contenant plusieurs gènes (duplications en bloc).

Le génome d'*A. thaliana* aurait subi au cours de son évolution plusieurs duplications massives, expliquant sa forte proportion de gènes dupliqués (AGI, 2000). Selon la méthode d'analyse, une duplication unique du génome entier (Blanc *et al.*, 2000; Lynch & Conery, 2000) ou plusieurs duplications de larges fragments (Ku *et al.*, 2000; Vision *et al.*, 2000) ont été identifiées. Ces événements de duplication massifs sont supposés s'être produits à différentes époques entre 50 à 220 MA. D'après la position, l'enchaînement et l'ordre des gènes dans les blocs dupliqués, l'évolution du génome d'*A. thaliana* résulterait plutôt de réarrangements successifs (duplications, délétions et divergence) d'un génome ancestral de Brassicacées non dupliqué plutôt que d'un événement de tétraploïdisation unique (Henry *et al.*, 2002). Récemment, une ré-analyse des données de Vision *et al.* (2000) a confirmé l'existence d'un événement de duplication massive des gènes d'*A. thaliana* vers 70 - 90 MA. Cependant, il existe des blocs dupliqués beaucoup plus anciens (jusqu'à 200 MA), indiquant des événements de duplications antérieurs (Raes *et al.*, 2003b).

## 2. Les autres duplications

Les duplications de larges fragments ne sont pas les seuls mécanismes de duplication : il existe des duplications de gènes uniques. Les copies issues de la duplication se retrouvent souvent côte à côte dans le génome. On parle alors de duplication en tandem. Des duplications en tandem successives peuvent conduire à la formation de groupes d'une dizaine de gènes localisés au même locus dans le génome. C'est le cas par exemple de certains *snoRNA* (Brown *et al.*, 2001). Des régions entières du génome d'*A. thaliana* sont constituées de gènes dupliqués en tandem, par exemple les loci des ARN ribosomiques (Pruitt & Meyerowitz, 1986; Campell *et al.*, 1992; Barakat *et al.*, 2001) et ceux des gènes de résistance aux maladies (Meyers *et al.*, 1999; Young, 2000; Baumgarten *et al.*, 2003). L'importance de ce mécanisme n'est donc pas à négliger dans la dynamique structurale du génome à l'échelle du gène (Bancroft, 2001). Par la combinaison de duplications en tandem successives et de duplications de fragments chromosomiques plus larges, un nombre important de gènes peut être obtenu à partir d'un seul gène ancestral.

Les éléments transposables participent à la dynamique du génome et par conséquent à la duplication des gènes (Bennetzen, 2000; Moffat, 2000; Bennetzen, 2002). Lors du déplacement ou de la duplication d'un élément transposable, l'étape d'excision est relativement imprécise et entraîne la duplication d'une région plus ou moins grande autour de lui (Schwarz-Sommer, 1987). Par ailleurs, la recombinaison homologue entre deux transposons peut entraîner le déplacement de grands fragments d'ADN génomiques (Courage-Tebbe *et al.*, 1983).

## C. L'émergence des familles multigéniques

### 1. La formation d'une famille

Les duplications à différentes échelles du génome génèrent une redondance de l'information génétique. Sous la pression évolutive, la redondance initiale a tendance à diminuer au cours du temps, entraînant une modification du génome sur le plan structural et fonctionnel (Wendel, 2000). La séquence et la structure d'un gène sont les deux principaux paramètres qui déterminent son expression et la protéine qu'il code, c'est-à-dire sa fonction. Ainsi, l'évolution moléculaire d'un gène, *i.e.* la modification de sa séquence et de sa structure, a pour conséquence son évolution fonctionnelle.

Lorsqu'un gène subit plusieurs événements de duplication successifs, suivis de la divergence des copies, cela aboutit à la formation d'une famille de gènes, c'est-à-dire un ensemble de gènes ayant le même ancêtre commun. Les gènes d'une famille ayant divergés au sein d'un même génome sont appelés paralogues, comme par exemple les gènes de globines  $\alpha$  et  $\beta$ . Les gènes ayant évolués après spéciation dans des organismes différents, par exemple les gènes de globines  $\alpha$  de souris et de lapin, sont appelés orthologues. L'hypothèse classique pose que les fonctions des gènes orthologues sont conservées d'un organisme à l'autre.

### 2. Les mécanismes d'évolution d'une famille

#### a. La sélection positive

Après la duplication d'un gène, la divergence des copies se produit par l'accumulation aléatoire de mutations dans la séquence. Si ces mutations sont délétères, le gène est inactivé. Il persiste dans le génome sous forme de pseudogène. Si ces mutations ont un effet positif, une nouvelle fonction apparaît dans l'une des copies. C'est la néofonctionnalisation (Nadeau & Sankoff, 1997). La nouvelle fonction est alors maintenue dans le génome par la sélection positive de la mutation bénéfique (Ohno, 1970). Cependant, ce modèle proposé par Ohno (1970) ne permet pas d'expliquer le maintien d'un fort pourcentage de gènes dupliqués au

sein des génomes eucaryotes. De plus, il suppose que l'une des copies du gène dupliqué est libre de toute pression de sélection. Or, il a été montré que le plus souvent les 2 copies sont soumises à une pression de sélection similaire (Hughes & Hughes, 1993).

#### b. Le modèle duplication-divergence-complémentation

Suite à ces observations, un nouveau modèle a été proposé (Force *et al.*, 1999). Selon ce modèle, l'accumulation de mutations délétères dans les deux copies provoque des pertes de fonctions partielles. La pression sélective étant dirigée sur les fonctions du gène ancestral, ces fonctions se répartissent entre les deux copies par un mécanisme de sous-fonctionalisation (Lynch & Force, 2000; Lynch *et al.*, 2001) et les deux copies deviennent complémentaires. Dans ce cas, la pression évolutive sur les deux copies est similaire. Il n'y a pas de sélection positive de nouvelles fonctions. Ce modèle duplication-divergence-complémentation (DDC) intègre l'hypothèse proposée par Ohno. En effet, dans le cas particulier où le gène dupliqué ne possède qu'une seule sous-fonction, l'accumulation des mutations conduira fréquemment à l'inactivation d'une des copies et parfois à l'apparition d'une nouvelle fonction. La répartition des sous-fonctions peut se réaliser par la modification de l'activité métabolique des protéines (Regenberg *et al.*, 1999; Zhang *et al.*, 2002), du niveau d'expression (Long & Dawid, 1980) ou de la localisation de l'expression des gènes (Ekker *et al.*, 1992; Mena *et al.*, 1996; Zhang & Choi, 2001). La répartition du profil d'expression peut également concerner différents stades de développement (Barry *et al.*, 1996).

Chez *Saccharomyces cerevisiae*, l'évolution d'éléments *cis*-régulateurs présents dans les gènes dupliqués résulte de la combinaison de pression de sélection positive et d'évolution selon le modèle DDC (Papp *et al.*, 2003). Les deux mécanismes, sélection positive et complémentation, co-existent donc dans le génome.

### 3. Conséquences fonctionnelles : redondance ou diversité ?

Les nombreuses duplications s'étant produites au sein du génome d'*A. thaliana* ont favorisé l'émergence de familles multigéniques de taille importante allant jusqu'à plusieurs centaines de gènes (AtPCMP, Aubourg *et al.*, 2000; P450, Paquette *et al.*, 2000). La présence de multiples copies à l'intérieur des familles de gènes soulève la question de savoir dans quelle mesure chaque membre de la famille pris individuellement est redondant en terme de fonction. En effet, un membre peut être considéré redondant s'il assure la même fonction que les autres membres de la famille ou non-redondant par un rôle spécifique qui lui est propre.

L'une des hypothèses avancées pour expliquer le maintien de gènes à l'état dupliqué dans le génome serait l'acquisition d'avantages subtils conférés par la duplication (Cooke *et al.*, 1997). Par exemple, les gènes ribosomiaux, redondants sur le plan fonctionnel, seraient conservés en plusieurs copies pour assurer le maintien d'une quantité suffisante d'ARNr (Long & Dawid, 1980). L'action synergique des gènes dupliqués pourrait être un autre de ces avantages subtils. Ce serait le cas des cyclines G1 au cours du cycle cellulaire chez *S. cerevisiae* (Nasmyth *et al.*, 1991). L'existence de gènes dupliqués impliqués dans les mécanismes cellulaires de base tels que la réplication, la transcription, l'épissage des ARNm et la traduction ne peut se concevoir que si cela ait un avantage à très long terme (Thomas, 1993).

Bien que des fonctions redondantes aient été observées au sein des familles multigéniques, de nombreuses études de mutants d'*A. thaliana* ont montré le rôle unique de certains membres. Sous leur apparente redondance, la fonction des membres d'une famille multigénique aurait donc tendance à diverger. L'acquisition progressive de différences, comme la modification de leur séquence et de leur structure, aboutissant au long terme à une différence fonctionnelle, permettrait le maintien de gènes dupliqués au sein du génome.

Les introns et les régions non codantes (UTR) participent aussi à l'évolution fonctionnelle des gènes. D'une part, la perte, le gain ou le déplacement d'introns et les modifications de la séquence des introns et des UTR permettent d'augmenter la proportion de régions divergentes entre les paralogues. En limitant ainsi les phénomènes de recombinaison homologues, des gènes peuvent être conservés plus longtemps à l'état dupliqué, favorisant l'acquisition de nouvelles fonctions. D'autre part, ces régions non codantes peuvent contenir des informations contrôlant l'expression des gènes. L'absence d'un intron peut avoir une conséquence directe sur la fonction d'un gène. Par exemple, chez la Pomme de terre, certains introns de gènes codant des saccharose synthases contrôlent le niveau et la répartition de l'expression de ces gènes (Fu *et al.*, 1995a; Fu *et al.*, 1995b). Chez l'Épinard, la régulation de l'expression du gène *PsaD* par la lumière dépend d'éléments structuraux présents dans ses introns (Bolle *et al.*, 1996). L'expression des gènes *p1* et *p2* chez le Maïs (homologues de facteurs MYB) dans des tissus différents est due à un changement de 5' UTR (Zhang *et al.*, 2000). Les sites de poly-adénylation des transcrits peuvent également influencer la spécificité tissulaire de l'expression (Edwalds-Gilbert *et al.*, 1997). Ainsi, les divergences de structure favorisent le maintien des gènes à l'état dupliqués et l'apparition de nouvelles fonctions.



## II. Problématique

### A. La régulation de l'expression comme moteur de l'évolution fonctionnelle

Immédiatement après la duplication d'un gène, les gènes dupliqués sont identiques et présentent des fonctions redondantes. Les mécanismes contrôlant la divergence des deux copies ne sont pas bien connus. Les deux séquences peuvent évoluer à la même vitesse ou non. Leur évolution peut concerner la séquence codante, les parties non codantes ou les régions régulatrices du gène dupliqué (Lynch & Force, 2000; Lynch *et al.*, 2001). L'organisation modulaire des éléments *cis*-régulateurs des promoteurs permettrait de changer radicalement l'expression des deux copies et donc d'introduire rapidement une divergence fonctionnelle entre elles. La modification de la régulation des gènes dupliqués formerait ainsi un module majeur de leur évolution fonctionnelle.

Plusieurs exemples montrent qu'un changement d'expression suffit à conférer une nouvelle fonction. Chez les animaux, les gènes de cristallines sont des enzymes du métabolisme recrutées dans le cristallin pour leurs propriétés optiques (Piatigorsky, 2003). Chez *A. thaliana*, les facteurs de transcription WEREWOLF et GLABRA1 de la famille MYB sont impliqués respectivement dans le contrôle de la formation des poils racinaires et des trichomes. Ces protéines sont équivalentes sur le plan fonctionnel, comme le montrent des expériences de complémentation réciproques. En effet, si l'une est exprimée à la place de l'autre, le développement de ces organes est normal. La spécificité de ces deux facteurs de transcription provient de leurs interactions avec des partenaires différents exprimés soit dans les feuilles, soit dans les racines (Kellogg, 2001).

Si la divergence fonctionnelle est l'un des mécanismes principaux du maintien des gènes dupliqués et, si une différence d'expression transcriptionnelle est un moyen d'introduire une divergence fonctionnelle, alors l'évolution de la régulation de l'expression des gènes est un point clé du contrôle de la dynamique du génome (Ferris & Whitt, 1979).

### B. Intérêts de l'étude de l'expression transcriptionnelle d'une famille multigénique

Comme un profil d'expression différent est une marque de spécialisation, de nombreux gènes dupliqués devraient présenter une divergence d'expression. Chez *S. cerevisiae*, des études de l'expression à l'échelle du génome ont montré que l'expression des gènes récemment dupliqués a tendance à diverger rapidement (Ferea *et al.*, 1999; Gu *et al.*, 2002). De plus, cette divergence semble avoir un rôle fonctionnel car elle est maintenue entre orthologues. En effet, les gènes appartenant à la même voie métabolique chez *S. cerevisiae* et *C. elegans* ont tendance à présenter le même profil d'expression (van Noort *et al.*, 2003). Associés à

l'expression transcriptionnelle, d'autres paramètres, comme la structure et la localisation des protéines, apportent des informations supplémentaires précieuses dans les études de l'évolution fonctionnelle des gènes à l'échelle du génome (Gerstein & Jansen, 2000). De même, l'étude de l'expression transcriptionnelle des gènes d'une famille en relation avec leur structure peut apporter des pistes quant à l'apparition de nouvelles fonctions au sein de cette famille.

Pour étudier l'évolution de la régulation de l'expression des gènes chez *A. thaliana*, il nous a semblé judicieux de nous placer dans le cadre d'une famille multigénique. D'une part, l'étude de l'évolution des familles donne accès à une partie de l'évolution du génome. En effet, les duplications massives qui ont conduit à l'expansion des familles au sein des génomes végétaux pourraient être corrélées avec les étapes majeures de l'évolution des plantes comme l'apparition des angiospermes ou la séparation des monocotylédones et des dicotylédones. Des traces de ces événements dans les arbres des familles multigéniques semblent confirmer cette hypothèse. Par exemple, chez *A. thaliana*, les actines sont réparties en deux groupes sur la base de leur proximité de séquence. Leur profil d'expression est corrélé avec ces groupes : l'un regroupe les actines s'exprimant dans les organes végétatifs, l'autre les actines spécifiques des organes reproducteurs (Meagher *et al.*, 1999). Une idée similaire est développée à partir de l'analyse de l'évolution fonctionnelle des facteurs de transcription MADS-box chez les plantes. L'apparition et la diversification de ces facteurs pourrait correspondre à l'apparition des structures reproductives comme l'ovule et la fleur (Theissen *et al.*, 2000). Ainsi, l'étude de l'évolution fonctionnelle d'une famille multigénique constitue une approche intéressante pour comprendre l'évolution du génome. La famille choisie comme modèle doit (1) comporter un nombre suffisant de gènes, (2) présenter des situations évolutives différentes et (3) être bien caractérisée sur le plan structural et fonctionnel. La famille des ARN hélicase à boîte DEAD (AtRH) étudiée au laboratoire a été retenue car elle respecte ces contraintes. Elle est présentée dans la partie suivante de l'introduction.

D'autre part, l'analyse d'une famille fournit un cadre intéressant pour mettre en place et développer une méthode de mesure de l'expression transcriptionnelle des gènes fiable, sensible et spécifique. En effet, cela permet de mettre au point la méthode sur un nombre de gènes réduit et dans une situation de similarité de séquences maximale. Une présentation des principales techniques de mesure de l'expression des gènes à grande échelle est développée en dernière partie de l'introduction. La forte similitude des séquences au sein d'une famille nous a incités à nous intéresser à la PCR quantitative en temps réel.

### C. Objectifs

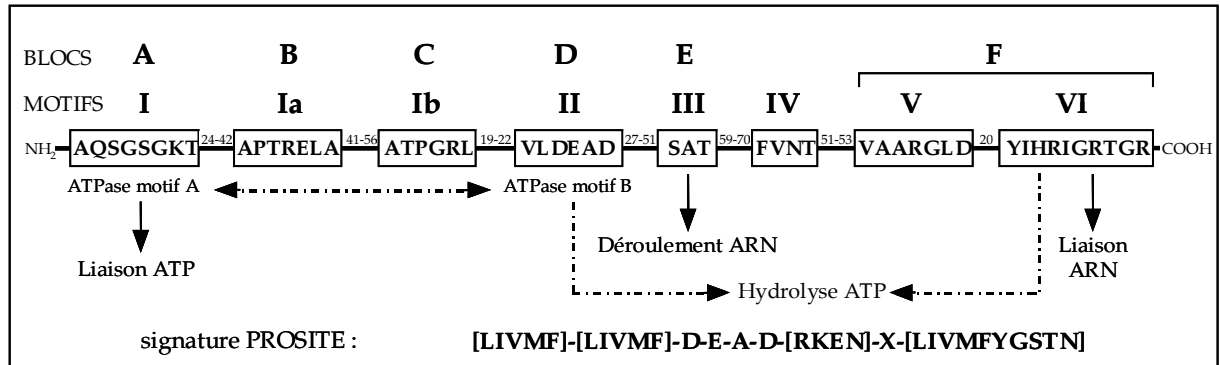
Les objectifs de ma thèse sont doubles : il s'agit, d'une part, de mesurer et d'analyser l'expression de la famille des *AtRH* d'un point de vue fondamental pour en tirer des conclusions quant au rôle de la régulation de l'expression dans l'évolution de cette famille et, d'autre part, d'évaluer la pertinence et l'efficacité de la PCR quantitative en temps réel pour la mesure de l'expression des gènes dans l'optique de l'appliquer ensuite au cadre d'une approche génomique.

### III. Les ARN hélicases à boîte DEAD : un modèle d'étude

Les ARN constituent un maillon central de l'expression des gènes. Les ARN messagers sont traduits en protéines par les ribosomes grâce au concours des ARN ribosomiques et des ARN de transfert. Par ailleurs, les *snRNA* ou petits ARN nucléaires participent au fonctionnement de la machinerie d'épissage des ARNm. Les *snoRNA* ou petits ARN nucléolaires guident la modification des ARNr lors de la formation des ribosomes (Bachelier & Cavaille, 1997; Maden & Hughes, 1997; Tollervey & Kiss, 1997; Filipowicz *et al.*, 1999). Certains ARN non codants sont impliqués dans d'autres fonctions comme la régulation de l'expression des gènes (Meli *et al.*, 2001). L'expression du génome dépend donc d'un grand nombre d'interactions ARN-ARN, ARN-protéine et ARN-ADN qui nécessitent à la fois le repliement correct des ARN et la réversibilité de ces interactions. En effet, la fonction des ARN est dépendante de la formation et/ou de l'élimination de structures secondaires c'est-à-dire du repliement et/ou du déploiement précis de leur séquence. Ces modifications de la structure tridimensionnelle des ARN font intervenir des molécules chaperonnes auxquelles participent pour une grande part les ARN hélicases.

#### A. Présentation de la famille

La fonction biochimique des ARN hélicases est de séparer les brins d'ARN lorsque deux brins d'ARN sont appariés entre eux (dans une structure en « épingle à cheveux » par exemple) ou lorsqu'un brin d'ARN est apparié à un brin d'ADN (lors de la transcription par exemple) en utilisant l'énergie libérée par l'hydrolyse de nucléosides triphosphates, l'ATP en général (Schmid & Linder, 1992; Lüking *et al.*, 1998; Linder *et al.*, 2001). Les ARN hélicases sont donc impliquées dans tous les processus du métabolisme des ARN : maturation des ARNr et biogenèse des ribosomes, transcription, épissage, édition et export des ARNm vers le cytoplasme, traduction et dégradation (Schmid & Linder, 1992; Lüking *et al.*, 1998; Linder *et al.*, 2001; Tanner & Linder, 2001). Certaines joueraient également un rôle dans les interactions ARN-protéine (Tanner & Linder, 2001).



**Figure 1 Organisation et fonctions catalytiques des résidus conservés des ARN hélicases à boîte DEAD** (d'après Aubourg, 1999).

Les relations entre les 8 motifs caractéristiques des ARN hélicases (Gorbalenya *et al.*, 1989) et les 6 blocs définis dans la base de données BLOCKS (Henikoff & Henikoff, 1991) pour les ARN hélicases à boîte DEAD sont indiqués. Les interactions physiques entre les motifs au niveau de la structure tertiaire sont représentées en pointillés. Les nombres entre les motifs indiquent le nombre d'acides aminés séparant classiquement les motifs. La longueur des extensions aux extrémités amino- et carboxy-terminales, respectivement avant le premier motif et après le dernier, varient de 20 à 400 aa.

Les ARN hélicases sont retrouvées dans tous les organismes vivants : des bactéries et des virus jusqu'à l'Homme en passant par les levures et les plantes (Gorbalenya *et al.*, 1989; Schmid & Linder, 1992; Lüking *et al.*, 1998; Linder *et al.*, 2001). Elles sont caractérisées par la présence de 7 à 8 groupes d'acides aminés conservés le long de leur séquence (Figure 1). Ces acides aminés correspondent à des résidus importants pour la fonction de l'hélicase comme les sites de liaison et d'hydrolyse de l'ATP et d'interaction avec l'ARN. Une classification a été établie sur la base de ces conservations. Il existe cinq grandes familles d'ARN hélicases non virales : les DEAD, les DEAH, les NS3-like, les Ski2-like et les Upf1-like (Tanner & Linder, 2001). Le nom des deux premières familles provient des quatre acides aminés conservés Asp-Glu-Ala-Asp ou Asp-Glu-Ala-His formant le quatrième motif caractéristique de ces protéines, aussi appelé motif II ou motif de Walker B, impliqué dans l'hydrolyse de l'ATP. Elles constituent la partie la plus importante des hélicases DExD/H (ou x peut être n'importe quel acide aminé) (Schmid & Linder, 1992; Lüking *et al.*, 1998; Aubourg *et al.*, 1999; Linder *et al.*, 2001; Tanner & Linder, 2001). Les trois autres familles tirent leur nom de celui de leur membre le plus représentatif.

La famille des ARN hélicases à boîte DEAD a été définie par Linder *et al.* (1989) chez *S. cerevisiae*. Le membre le plus connu et le plus étudié est le facteur d'initiation de la traduction eIF4A (Linder & Slonimski, 1989; Prat *et al.*, 1990; Schmid & Linder, 1991; Metz *et al.*, 1992; Pause & Sonenberg, 1992; Pause *et al.*, 1993; Brander *et al.*, 1995; Mandel *et al.*, 1995; Li *et al.*, 2001b; Rogers *et al.*, 2001a; Rogers *et al.*, 2001b; Svitkin *et al.*, 2001; Tanner *et al.*, 2003). Chez les plantes, les différentes études réalisées sur des ARN hélicases à boîte DEAD sont rassemblées ci-dessous (Tableau 1).

**Tableau 1 Repères historiques des ARN hélicases à boîte DEAD chez les plantes.**

Année	Découvertes	Références
1977	Première mention d'une activité ARN hélicase au cours de l'initiation de la transcription dans des réticulocytes de Lapin	(Ilan, 1977)
1985	Etude de l'activité ARN hélicase d'eIF4A de Lapin	(Ray <i>et al.</i> , 1985)
1989	Définition de la « famille des ARN hélicases à boîte DEAD »	(Linder <i>et al.</i> , 1989)
1990	Première mesure d'une activité ARN hélicase chez les plantes : celle du facteur eIF4A extrait du germe de Blé	(Jaramillo <i>et al.</i> , 1990)
1992	Premières séquences nucléiques d'ARN hélicases à boîte DEAD d' <i>A. thaliana</i> : celles de deux ADNc codant des facteurs d'initiation de la transcription eIF4A (At3g13920 ; At1g54270)	(Metz <i>et al.</i> , 1992)
1994	Structure et expression d'un ADNc codant une ARN hélicase chez le Tabac	(Itadani <i>et al.</i> , 1994)
1995	Identification d'une ARN hélicase à boîte DEAD spécifique du pollen et de plusieurs autres ARN hélicases à boîte DEAD chez le Tabac Expression constitutive du gène eIF4A chez <i>A. thaliana</i>	(Brander & Kuhlemeier, 1995; Brander <i>et al.</i> , 1995) (Mandel <i>et al.</i> , 1995)
1997	Identification du gène <i>AtRHI</i> (At4g15850) et de 18 autres gènes codant des ARN hélicases à boîte DEAD chez <i>A. thaliana</i> Identification de PRH75 (At5g62190), une ARN hélicase à boîte DEAD localisée dans le noyau chez <i>A. thaliana</i> et l'Épinard	(Aubourg <i>et al.</i> , 1997) (Lorkovic <i>et al.</i> , 1997)
1998	Caractérisation d'une ATPase/ARN hélicase à boîte DEAD d' <i>A. thaliana</i> (At3g01540)	(Okanami <i>et al.</i> , 1998)
1999	Caractérisation de la famille AtRH des ARN hélicases à boîte DEAD d' <i>A. thaliana</i> Identification d'une ARN hélicase à boîte DEAD régulée au cours du mûrissement chez la Tomate	(Aubourg <i>et al.</i> , 1999) (Zegzouti <i>et al.</i> , 1999)
2000	Caractérisation du gène VDL codant une ARN hélicase à boîte DEAD chloroplastique chez le Tabac	(Wang <i>et al.</i> , 2000b)
2001	Evolution de la structure intron/exon des gènes de la famille des ARN hélicases à boîte DEAD chez <i>A. thaliana</i> , <i>C. elegans</i> et <i>D. melanogaster</i> Caractérisation d'une ARN hélicase à boîte DEAD de Haricot mungo	(Boudet <i>et al.</i> , 2001) (Li <i>et al.</i> , 2001a)
2002	Implication d'une ARN hélicase à boîte DEAD d' <i>A. thaliana</i> (At3g53110) dans la résistance au froid et au gel	(Gong <i>et al.</i> , 2002)

### B. Modèle d'étude de l'évolution de la régulation de l'expression transcriptionnelle

La famille des ARN hélicases à boîte DEAD d'*A. thaliana* a été caractérisée au laboratoire par Sébastien Aubourg et Nathalie Boudet (Aubourg *et al.*, 1997; Aubourg *et al.*, 1999; Boudet *et al.*, 2001). Cette famille compte 1 à 5 gènes dans les génomes bactériens entièrement séquencés, 26 gènes chez la levure *S. cerevisiae*, 32 et 30 gènes chez *C. elegans* et *D. melanogaster*. Chez *A. thaliana*, cette famille comporte 58 paralogues, avec des gènes récemment dupliqués et des gènes issus de duplications plus anciennes (Boudet *et al.*, 2001). Cette augmentation du nombre de membres de la famille des ARN hélicases à boîte DEAD reflète une caractéristique des génomes végétaux, à savoir présenter un plus grand nombre de gènes dupliqués que les génomes animaux (AGI, 2000). Cette augmentation est également associée à une taille plus importante de toutes les familles impliquées dans le métabolisme des ARN chez *A. thaliana* par rapport à celle des autres eucaryotes (Anantharaman *et al.*, 2002). La structure intron/exon des gènes *AtRH* est très variable : de 0 à 18 introns (Boudet *et al.*, 2001). Cette importante divergence structurale pourrait avoir un rôle biologique. Plusieurs études suggèrent l'expression spécifique de certaines hélicases à boîte DEAD suite à des stress, au cours du développement ou au cours du cycle cellulaire chez les animaux (De Valoir *et al.*, 1991; Roussell & Bennett, 1993; Fujiwara *et al.*, 1994; Eberl *et al.*, 1997; Chamot *et al.*, 1999; Fujimura & Takamura, 2000; MacArthur *et al.*, 2000; Navarro *et al.*, 2001) et chez les plantes (Brander & Kuhlemeier, 1995; Zegzouti *et al.*, 1999). Mais, aucune étude exhaustive de l'expression des ARN hélicases à boîte DEAD n'a été réalisée chez un eucaryote pluricellulaire.

Par le nombre de paralogues, la diversité des séquences, celle de la structure des gènes et l'existence potentielle de spécialisations fonctionnelles, cette famille présente une situation évolutive intéressante, correspondant à notre problématique.



#### **IV. De l'étude du transcriptome à l'étude de l'expression transcriptionnelle d'une famille**

Le génome, ensemble du patrimoine génétique d'un individu, a été longtemps considéré comme un ensemble statique d'informations : la séquence d'ADN de ses chromosomes. La dynamique cellulaire est le reflet de l'expression concertée du génome. Le transcriptome, ensemble des transcrits d'une cellule, constitue un ensemble sans cesse en adaptation, résultant des réactions permanentes de la cellule aux informations qu'elle perçoit. Avoir accès aux modulations de l'expression du génome en fonction de la physiologie cellulaire ouvre de nouvelles perspectives dans la compréhension du fonctionnement de la cellule. De récentes approches ont rendu possible l'étude simultanée de l'expression d'un grand nombre de gènes. Le regroupement de ces gènes par profils d'expression similaires permet alors la caractérisation de réseaux de gènes co-régulés (Ruan *et al.*, 1998; Chen *et al.*, 2002). Les approches de l'analyse du transcriptome dites « à grande échelle » mettent en oeuvre différentes techniques. Les programmes de séquençage d'EST (*Expressed Sequence Tag*), d'étiquettes SAGE (*Serial Analysis of Gene Expression*) et la technique MPSS (*Massively Parallel Signature Sequencing*) sont basés sur le recensement des transcrits via leur étiquetage. D'autres approches exploitent les variations de taille de fragments d'ADN pour distinguer les différents transcrits, par exemple la technique de ADNc-AFLP (*ADNc-Amplified Fragment Length Polymorphism*). Enfin, les « puces à ADN » ou encore *microarrays*, utilisent l'hybridation de sondes nucléiques marquées à des transcrits, en général partiels, immobilisés sur un support solide.

##### A. Les moyens d'étude de l'expression transcriptionnelle à grande échelle

###### 1. L'étiquetage des transcrits

###### a. Les EST, étiquettes de séquences exprimées

La première analyse de l'expression d'un grand nombre de gènes a pris la forme de programmes de séquençage de transcrits à grande échelle. Des banques de transcrits ont été constituées à partir d'ARNm provenant de matériels biologiques variés et de diverses conditions physiologiques. La majorité de ces programmes se sont attachés à séquencer les extrémités de clones d'ADNc afin d'étiqueter un maximum de transcrits. Les séquences obtenues, appelées EST, sont de courtes séquences de 300 à 500 nucléotides, issues du séquençage unique de l'extrémité 5' ou 3' d'un ADNc. La qualité de ces séquences est relativement mauvaise car une EST peut comporter jusqu'à 5% d'erreurs et d'incertitudes. Cependant, elle est largement suffisante pour identifier avec certitude l'ADNc étiqueté. Ainsi,

les EST fournissent un accès direct à la partie exprimée du génome. Cette approche a été largement utilisée lors de l'analyse de différents génomes comme celui du nématode *C. elegans* (Waterston *et al.*, 1992), de la drosophile *D. melanogaster* (Rubin *et al.*, 2000; Stapleton *et al.*, 2002a; Stapleton *et al.*, 2002b) ou de l'Homme (Adams *et al.*, 1991; Adams *et al.*, 1992; Adams *et al.*, 1993; Adams *et al.*, 1995; Hillier *et al.*, 1996; Ewing & Green, 2000). Chez les plantes, elle a permis de caractériser plusieurs milliers de gènes chez *A. thaliana* (Höfte *et al.*, 1993; Newman *et al.*, 1994; Cooke *et al.*, 1996; Asamizu *et al.*, 2000; White *et al.*, 2000), le Riz (Yamamoto & Sasaki, 1997; Ewing *et al.*, 1999; Wang *et al.*, 2001; De Los Reyes *et al.*, 2003) et d'autres espèces cultivées (Maïs, Sorgho, Cacao, Avoine...).

Très utiles pour déterminer les zones transcrites d'un gène et caractériser sa structure intron/exon, les EST apportent également des indications qualitatives et quantitatives sur sa transcription. En effet, lorsqu'un nombre suffisant d'EST est obtenu pour une banque donnée, certaines EST correspondent au même gène, indiquant ainsi la présence dans la banque de plusieurs ADNc associés à ce gène. Ainsi, plus un gène est exprimé, plus le nombre d'EST correspondant à ce gène dans la banque est élevé. Les ADNc étiquetés sont donc représentatifs des gènes les plus exprimés dans un organe donné et dans une situation physiologique précise. Lorsqu'un gène est identifié dans une seule banque d'ADNc, on peut penser qu'il s'exprime spécifiquement dans l'échantillon de matériel végétal utilisé pour générer la banque. En multipliant l'origine des ARNm utilisés pour générer les banques d'ADNc, le nombre de situations physiologiques analysables augmente.

Si les EST sont bien représentatives de la banque d'ADNc dont elles sont issues, elles peuvent permettre d'estimer l'expression relative des gènes, en rapportant le nombre d'EST obtenu pour chaque gène au nombre total d'EST. Ainsi, un profil d'expression des gènes peut être obtenu en effectuant ce calcul sur plusieurs banques d'ADNc représentant les transcrits de différents matériels végétaux. Plus le nombre d'EST est important, plus l'estimation de l'expression des gènes est fiable. La qualité des analyses de l'expression du génome à partir des EST est donc directement liée au nombre de séquences (Audic & Claverie, 1997). En novembre 2003 (release 111403), le nombre d'EST dans la base de données dbEST était de 5.427.257 pour l'Homme et de 190.732 pour *A. thaliana*.

L'abondance des EST peut être utilisée pour la caractérisation fonctionnelle de familles de gènes. L'identification des familles des gènes impliquées dans la biosynthèse de la lignine en est un exemple d'utilisation fructueuse (Raes *et al.*, 2003a). 160.776 EST d'*A. thaliana* ont

été réparties selon leur origine en 11 catégories, comprenant différents organes et différents stress. Les EST correspondant aux 34 gènes de l'étude ont été identifiées, puis leur abondance relative dans les différentes catégories a été établie. Dans l'ensemble, l'abondance relative des EST a fourni des informations similaires aux données de RT-PCR semi-quantitative pour ces 34 gènes, même si des incohérences et des contradictions peuvent être relevées (Raes *et al.*, 2003a). Les EST d'*A. thaliana* ont été également utilisées pour établir le profil transcriptionnel des AtPEM, protéines de la membrane interne des plastes, à partir de 55 banques d'ADNc réparties en 8 classes « tissus-spécifiques » (Koo & Ohlrogge, 2002).

Le développement de nouvelles méthodes de sélection des ARNm entiers a permis de compléter les approches EST par des programmes de séquençage de transcrits pleine longueur, par exemple en utilisant la coiffe des ARNm comme point d'ancrage au lieu de la queue poly-A (Seki *et al.*, 1998; Seki *et al.*, 2002c; Seki *et al.*, 2002d; Wu *et al.*, 2002). S'ils apportent des données précieuses pour l'annotation des gènes (Castelli *et al.*, 2004), ils ne fournissent pas plus d'indications quantitatives sur l'expression des gènes que les EST.

#### b. Les étiquettes SAGE

La technique SAGE consiste également en l'étiquetage des transcrits mais avec des séquences plus courtes. Des étiquettes d'une dizaine de nucléotides sont donc générées par digestion enzymatique à partir de l'extrémité 3' des ARNm. Leur séquençage après concaténation permet de déterminer leur abondance relative (Velculescu *et al.*, 1995; Velculescu *et al.*, 1997; Wang & Rowley, 1998; Velculescu *et al.*, 2000).

L'inconvénient majeur est la taille des étiquettes, souvent trop courtes pour identifier sans ambiguïté les transcrits (Lee *et al.*, 2002). Des améliorations de la technique ont été développées pour palier à ce problème (Chen *et al.*, 2003), réduire la quantité d'ARN nécessaires (Datson *et al.*, 1999) ou l'efficacité de la construction de la banque (Powell, 1998; Kenzelmann & Muhlemann, 1999; Angelastro *et al.*, 2000; Du *et al.*, 2003).

Chez les plantes, des profils d'expression ont été recherchés dans différents organes et différentes conditions physiologiques. L'expression des gènes dans les feuilles et les graines (Gibbins *et al.*, 2003), des gènes induits par un éliciteur du pathogène *Magnaporthe grisea* (Matsumura *et al.*, 2003) et des gènes impliqués dans la croissance en anaérobie (Matsumura *et al.*, 1999) a été analysée chez le Riz. Chez *A. thaliana*, la technique SAGE a été utilisée pour étudier la modulation de l'expression des gènes en réponse au froid (Jung *et al.*, 2003; Lee & Lee, 2003) et dans les cellules racinaires en fonction de la source d'azote (Fizames *et*

*al.*, 2004). Enfin, cette méthode a permis de caractériser des gènes impliqués dans la formation du bois chez le pin *Pinus taeda* (Lorenz & Dean, 2002).

#### c. La technique MPSS

Le séquençage massif d'étiquettes de transcrits MPSS s'inspire très largement de la technique SAGE. Il s'agit en effet de dénombrer des étiquettes de 17 nucléotides obtenues après digestion enzymatique de l'extrémité 3' des ARNm. La particularité de cette technique réside dans la méthode de séquençage employée. Les étiquettes sont fixées à des billes immobilisées sur une surface solide, chaque bille portant les mêmes étiquettes. Elles sont ensuite séquencées en parallèle via un système à quatre fluorochromes. A chaque cycle de séquençage, les billes sont marquées du fluorochrome correspondant au nucléotide séquencé. La fluorescence des billes est enregistrée à chaque cycle en quatre couleurs, générant une image proche de celles des puces à ADN. L'analyse des images successives permet de déterminer en parallèle la séquence des étiquettes fixées à chaque bille (Brenner *et al.*, 2000a). Un site internet donne l'accès aux résultats obtenus pour le génome d'*A. thaliana* aux adresses : <http://dbixs001.dbi.udel.edu/MPSS4/java.html> ou <http://mpss.udel.edu/at/>.

Une étude du mutant d'*A. thaliana* insensible à l'acide abscissique *abi1-1* avec cette technique a permis d'identifier de nouveaux gènes régulés par cette hormone (Hoth *et al.*, 2002).

#### d. Conclusions

Le principe de compter les étiquettes est une méthode d'analyse de l'expression directe mais elle nécessite de nombreuses manipulations. La fiabilité des résultats est liée à la qualité des banques générées. Cependant, cette méthode est plus performante pour les gènes fortement exprimés. Pour étudier les gènes de faible expression, il faut augmenter la taille des banques et le nombre de séquences réalisées. C'est d'ailleurs cette raison qui a motivé l'arrêt des premiers programmes de séquençage systématique d'EST d'*A. thaliana* car pour obtenir la séquence d'un nouveau transcrit, il aurait fallu séquencer un trop grand nombre de clones correspondant à des transcrits déjà connus et étiquetés. Actuellement, de nombreuses EST continuent d'être déposées dans les bases de données publiques. Elles proviennent de nouvelles banques qui permettent, par exemple de comparer l'expression entre deux conditions. La soustraction par hybridation permet d'éliminer les transcrits dont la teneur est similaire dans les deux conditions. Ainsi, la banque finale est enrichie en transcrits présentant une expression différente et appauvrie d'une partie des transcrits les plus abondants (Bonaldo *et al.*, 1996; Lombardi *et al.*, 2003).

Tableau 2 Quelques analyses à l'aide de puces à ADNc chez les plantes.

Référence	Espèce
(Girke <i>et al.</i> , 2000)	<i>A. thaliana</i>
(Schenk <i>et al.</i> , 2000)	<i>A. thaliana</i>
(Wang <i>et al.</i> , 2000a)	<i>A. thaliana</i>
(Perez Amador <i>et al.</i> , 2001)	<i>A. thaliana</i>
(Schaffer <i>et al.</i> , 2001)	<i>A. thaliana</i>
(Xu <i>et al.</i> , 2001)	<i>A. thaliana</i>
(Yu <i>et al.</i> , 2001)	<i>A. thaliana</i>
(Gutierrez <i>et al.</i> , 2002)	<i>A. thaliana</i>
(Kurth <i>et al.</i> , 2002)	<i>A. thaliana</i>
(Ramonell <i>et al.</i> , 2002)	<i>A. thaliana</i>
(Ruuska <i>et al.</i> , 2002)	<i>A. thaliana</i>
(Cartieaux <i>et al.</i> , 2003)	<i>A. thaliana</i>
(Folta <i>et al.</i> , 2003)	<i>A. thaliana</i>
(Golem & Culver, 2003)	<i>A. thaliana</i>
(Kotake <i>et al.</i> , 2003)	<i>A. thaliana</i>
(Krizek <i>et al.</i> , 2003)	<i>A. thaliana</i>
(Narusaka <i>et al.</i> , 2003)	<i>A. thaliana</i>
(Zik & Irish, 2003)	<i>A. thaliana</i>
(Horvath <i>et al.</i> , 2003b)	<i>Avena fatua</i> L. (Avoine)
(Horvath <i>et al.</i> , 2003a)	<i>Euphorbia esula</i> L., Peuplier, Avoine
(Aharoni & O'Connell, 2002)	<i>Fragaria x ananassa</i> Duchesne ex Rozier (Fraise)
(Maguire <i>et al.</i> , 2002)	<i>Glycine max</i> (L.) Merr. (Soja)
(Faccioli <i>et al.</i> , 2002)	<i>Hordeum vulgare</i> L. (Orge)
(Negishi <i>et al.</i> , 2002)	<i>Hordeum vulgare</i> L. (Orge)
(Zhao <i>et al.</i> , 2003)	<i>Lycopersicon esculentum</i> Mill. (Tomate)
(Akimoto-Tomiyama <i>et al.</i> , 2003)	<i>Oryza sativa</i> L. (Riz)
(Yazaki <i>et al.</i> , 2000)	<i>Oryza sativa</i> L. (Riz)
(Stasolla <i>et al.</i> , 2003)	<i>Picea glauca</i> (Moench) Voss (Epinette blanche)
(Yang <i>et al.</i> , 2003)	<i>Robinia pseudoacacia</i> L.
(Jones <i>et al.</i> , 2002)	<i>Theobroma cacao</i> L. (Cacao)
(Sawbridge <i>et al.</i> , 2003)	<i>Trifolium repens</i> L. (Trèfle)
(Nakazono <i>et al.</i> , 2003)	<i>Zea mays</i> L. (Maïs)
(Demura <i>et al.</i> , 2002)	<i>Zinnia elegans</i> Jacq.

## 2. ADNc-AFLP

Cette technique permet de caractériser le profil d'expression transcriptionnelle d'un ensemble de gènes à l'échelle du génome entier. Elle est basée sur le polymorphisme de la longueur de fragments d'ADN amplifiés par PCR. L'analyse de l'intensité des bandes dans les gels permet de quantifier l'expression. Chaque bande est identifiée d'abord par sa taille puis par séquençage ultérieur. Comme avec les étiquettes de transcrits, l'analyse est réalisée sans a priori. Cette technique a permis la caractérisation de gènes régulés au cours de la division cellulaire dans des cellules BY-2 de tabac synchronisées à l'aide de l'aphidicoline (Breyne *et al.*, 2003). Une revue compare la technique de ADNc-AFLP aux *microarrays* et leur pertinence respective pour établir le profil d'expression de l'ensemble des gènes d'*A. thaliana* (Donson *et al.*, 2002).

## 3. Les puces à ADN

L'étude de l'expression des gènes à l'aide des puces à ADN est basée sur l'hybridation de sondes nucléiques avec des fragments d'ADN immobilisés sur un support solide. Par leur miniaturisation, les *microarrays* constituent une avancée technologique impressionnante. Il est possible de déposer sur quelques mm<sup>2</sup> à quelques cm<sup>2</sup> plusieurs centaines à plusieurs milliers de séquences. Ainsi, il est possible de réaliser des *microarrays* correspondant à l'ensemble des gènes d'un génome (Lashkari *et al.*, 1997). Inventée et mise au point sur quelques dizaines de gènes d'*A. thaliana* (Schena *et al.*, 1995), cette technique a été appliquée depuis à d'autres génomes modèles comme ceux d'*Escherichia coli*, *S. cerevisiae*, *D. melanogaster*, *Mus musculus* et *Homo sapiens sapiens*. Actuellement, il existe plusieurs types de puces à ADN. Elles diffèrent par la nature des séquences immobilisées : il s'agit soit de fragments d'ADN de plusieurs centaines de bases (ADNc ou produits PCR), soit d'oligonucléotides, et par la façon de les fixer à la puce : par dépôt (ADNc, produits PCR et longs oligonucléotides) ou par synthèse *in situ* (oligonucléotides courts uniquement).

### a. Les puces d'ADNc ou de produits PCR

De nombreuses puces regroupent des ADNc plus ou moins partiels identifiés par une EST. Certaines puces sont basées sur des ADNc pleine longueur (Seki *et al.*, 2001; Seki *et al.*, 2002a; Seki *et al.*, 2002b; Oono *et al.*, 2003). Largement utilisées pour étudier l'expression des gènes d'*A. thaliana*, elles le sont également pour les espèces d'intérêt agronomique (Tableau 2). En effet, la connaissance du génome entier d'*A. thaliana* facilite l'annotation des

autres génomes végétaux par homologie et donc l'exploitation et la valorisation des résultats obtenus pour ces espèces.

Ces puces sont partielles : au mieux, la moitié du transcriptome (14.500 gènes environ) est représentée (Cartieaux *et al.*, 2003; Sawbridge *et al.*, 2003). Le nombre de gènes est plutôt de l'ordre de 7000 à 8000 en moyenne pour les puces dédiées à *A. thaliana*, contre 2000 à 3000 gènes (à peine 10% du transcriptome) pour les puces dédiées aux autres espèces végétales. Certaines puces spécialisées ne comportent que l'équivalent d'environ 2000 gènes d'*A. thaliana* impliqués dans une fonction donnée, par exemple les gènes de défense (Schenk *et al.*, 2000; Ramonell *et al.*, 2002) ou les gènes plastidiaux (Kurth *et al.*, 2002).

La puce CATMA est issue d'un projet de recherche européen<sup>1</sup> (Hilson *et al.*, 2004). Elle contient des fragments d'ADN de 150 à 500 pb amplifiés par PCR appelés GST (*Gene Sequence Tag*) représentant 23038 gènes<sup>2</sup> d'*A. thaliana* (Crowe *et al.*, 2003). Ces GST correspondent à la région la plus divergente de chaque gène. Il s'agit le plus souvent des régions 3' UTR ou 5' UTR (Thareau *et al.*, 2003). Ainsi, contrairement aux puces basées sur des ADNc clonées, la puce CATMA a été développée pour limiter au maximum les hybridations croisées des transcrits d'une même famille entre les séquences des différents paralogues fixées à la puce. Elle permet donc de mesurer l'expression des gènes avec une meilleure spécificité. Aucune étude du transcriptome d'*A. thaliana* avec la puce CATMA n'est publiée à ce jour mais, en France, au moins neuf projets de recherche utilisant cette puce sont en cours<sup>3</sup>.

Les puces à ADN ne comportent pas toutes des séquences de transcrits. Une partie de la séquence du chromosome V d'*A. thaliana* a été déposée sur un filtre à haute densité, à raison d'environ 3kb par clone d'ADN génomique. Cela a permis d'identifier des zones transcrites non connues (Hanano *et al.*, 2002). La séquence génomique d'un organisme peut également être déposée sous forme d'oligonucléotides. Ce type de puce a été utilisé pour identifier les zones exprimées du génome hors des gènes connus du chromosome 22 humain avec des oligonucléotides de 60-mers (Rinn *et al.*, 2003) et du génome d'*A. thaliana* avec 12 x 834000 oligonucléotides de 25-mers (Yamada *et al.*, 2003).

---

<sup>1</sup> <http://www.catma.org/>

<sup>2</sup> [http://www.evry.inra.fr/public/projects/transcriptome/production\\_fr.html](http://www.evry.inra.fr/public/projects/transcriptome/production_fr.html)

<sup>3</sup> [http://www.evry.inra.fr/public/projects/transcriptome/collaborations\\_genoplante.html](http://www.evry.inra.fr/public/projects/transcriptome/collaborations_genoplante.html)

### b. Les puces d'oligonucléotides

Deux modèles de puces basés sur des oligonucléotides synthétisés *in situ* (technologie Affymetrix) existent pour étudier l'expression du génome d'*A. thaliana*. La première puce GeneChip AG1 est composée d'oligonucléotides de 25-mers représentant environ de 8300 gènes d'*A. thaliana* (Zhu *et al.*, 2001). La nouvelle puce ATH1 contient 22.500 jeux d'oligonucléotides correspondant à environ 24.000 gènes (Wang *et al.*, 2003a).

Une vingtaine d'études basées sur ces puces à oligonucléotides chez *A. thaliana* ont été publiées. Elles portent sur la modification du transcriptome en réponse à des stress biotiques (Puthoff *et al.*, 2003; Tao *et al.*, 2003; Whitham *et al.*, 2003) ou abiotiques (Moseyko *et al.*, 2002; Lechelt-Kunze *et al.*, 2003; op den Camp *et al.*, 2003; Ulm *et al.*, 2004), selon le rythme circadien (Harmer *et al.*, 2000), selon le cycle cellulaire (Menges *et al.*, 2002), selon l'action de différentes hormones (Mussig *et al.*, 2002; Monroe-Augustus *et al.*, 2003; Mussig *et al.*, 2003; Ogawa *et al.*, 2003; Rashotte *et al.*, 2003), de mutants de développement (Tepperman *et al.*, 2001; Schroeder *et al.*, 2002; Rizhsky *et al.*, 2003; William *et al.*, 2004), pendant le développement du pollen (Becker *et al.*, 2003; Honys & Twell, 2003), pendant la germination (Duque & Chua, 2003), au cours du métabolisme (Hammond *et al.*, 2003; Laule *et al.*, 2003; Maathuis *et al.*, 2003; Wang *et al.*, 2003a) et lors de l'activation de la voie de dégradation des protéines mal repliées via le facteur de transcription *IRE1* (Martinez & Chrispeels, 2003; Noh *et al.*, 2003).

### c. Les puces représentant le génome entier

Disposer de l'ensemble du génome permet une meilleure analyse du transcriptome que dans le cas des puces partielles. Dans le tableau suivant (Tableau 3), est rassemblé l'ensemble des données issues de puces représentant le génome complet d'*A. thaliana* sous forme d'oligonucléotides (ATH1, Agilent oligo) ou de produits PCR (CATMA, TIGR). Les données de ces 66 expérimentations sont le plus souvent disponibles dans des bases de données publiques, accessibles par des sites internet dédiés. Certaines sont associées à une publication. On compte 36 analyses de mutants, 38 analyses de l'effet de traitements, 7 expériences de contrôle, 6 cinétiques et 3 comparaisons de tissus.

Les puces apportent une vision d'ensemble de la régulation du transcriptome en réponse à divers stimuli ou en fonction de différentes conditions physiologiques. C'est un outil intéressant pour caractériser des groupes de gènes co-régulé et ainsi identifier des éléments *cis*-régulateurs spécifiques présents dans les promoteurs de gènes de ces groupes (par exemple : Spellman *et al.*, 1998; Maleck *et al.*, 2000; Pufky *et al.*, 2003).



**Tableau 3 Données des puces représentant le génome complet d'*A. thaliana* publiées ou disponibles sur internet** (Vincent Thareau, communication personnelle).

Lorsque des données sont accessibles sur internet, le site est indiqué. La nature des séquences immobilisées varie selon la puce. La puce ATH1 est composée d'oligonucléotides de 25-mers synthétisés *in situ*. La puce d'Agilent oligo contient des oligonucléotides de 60-mers, la puce CATMA contient des fragments d'ADN de 150 à 500 pb et celle du TIGR des fragments d'ADN génomique de 1000 pb. Les auteurs ou les expérimentateurs, une brève description des analyses réalisées et le matériel végétal utilisé sont précisés dans les colonnes suivantes.

Site Internet	Puce	Auteurs	Expériences	Organes
4	ATH1	(op den Camp <i>et al.</i> , 2003)	Cinétique des effets de la lumière et du paraquat sur le mutant conditionnel « fluorescent » (flu)	feuilles de rosette
-	ATH1	(Duque & Chua, 2003)	Analyse du mutant du facteur de transcription <i>imbl</i> pendant la germination des graines	graines
-	ATH1	(Monroe-Augustus <i>et al.</i> , 2003)	Analyse du mutant de réponse à l'acide indole-3-butérique (auxine) <i>ibr5</i>	jeunes plantes
5	ATH1	(Mussig <i>et al.</i> , 2003)	Effet des brassinostéroïdes sur des mutants BR et des plantes de phénotype sauvage	racines
-	ATH1	(Ulm <i>et al.</i> , 2004)	Effet des rayonnements UV	jeunes plantes
6	ATH1	(Wang <i>et al.</i> , 2003a)	Effet du nitrate	racines et parties aériennes
7	ATH1	(William <i>et al.</i> , 2004)	Analyse des mutants d'identité du méristème <i>leafy</i> , <i>apetala1</i> et <i>cauliflowers</i>	jeunes plantes
Arrayexpress <sup>8</sup>	ATH1	inconnu	Effet des nématodes sur l'expression des gènes d' <i>A. thaliana</i>	racines latérales
Arrayexpress <sup>5</sup>	ATH1	Casal	Analyse de mutants de phytochrome et de cryptochrome	plante entière
Arrayexpress <sup>5</sup>	ATH1	Cornah	Analyse de deux mutants <i>knock-out</i> d'enzymes du cycle du glyoxylate : la malate synthase et l'isocitrate lyase	jeunes plantes
Arrayexpress <sup>5</sup>	ATH1	Greco	Analyse du mutant du facteur de remodelage de la chromatine BOU	feuilles de rosette
Arrayexpress <sup>5</sup>	ATH1	Millenaar	Induction de la croissance hyponastique par l'éthylène	pétiole
Arrayexpress <sup>5</sup>	ATH1	Murray	Effet de l'inoculation de <i>Pseudomonas syringae</i> sur l'expression constitutive de gènes de défense du mutant <i>cir1</i> et celle d'une plante de phénotype sauvage	feuilles
Arrayexpress <sup>5</sup>	ATH1	Rentel	Effet de l'H <sub>2</sub> O <sub>2</sub> sur le mutant OX1 et sur une plante de phénotype sauvage	jeunes plantes
Arrayexpress <sup>5</sup>	ATH1	Villadsen	Effet de différents sucres sur l'expression des gènes d' <i>A. thaliana</i>	jeunes plantes
Arrayexpress <sup>5</sup>	ATH1	Ward	Analyse de mutants hypersensibles ou résistants à l'auxine pour caractériser des gènes impliqués dans l'activation des méristèmes des bourgeons axillaires	bourgeons axillaires
Arrayexpress <sup>5</sup>	CATMA	Brenner	Comparaison des écotypes Colombia et Landsberg erecta	parties aériennes
Arrayexpress <sup>5</sup>	CATMA	Brenner	Comparaison de mêmes conditions de culture entre Ghent et Berlin	parties aériennes
Arrayexpress <sup>5</sup>	CATMA	Brenner	Comparaison de deux séries d'ADNc issues du même extrait d'ARN totaux	parties aériennes
Arrayexpress <sup>5</sup>	CATMA	Hadji	Comparaison des écotypes Colombia et Landsberg erecta	plante entière
Arrayexpress <sup>5</sup>	CATMA	Hadji	Comparaison des mêmes conditions de culture entre Ghent et Evry	plante entière
Arrayexpress <sup>5</sup>	CATMA	Hadji	Comparaison d'un échantillon d'ADNc d' <i>A. thaliana</i> avec lui-même	plante entière
Arrayexpress <sup>5</sup>	CATMA	Little	Comparaison des écotypes Colombia et Landsberg erecta	plante entière
Arrayexpress <sup>5</sup>	CATMA	Little	Comparaison des mêmes conditions de culture entre Ghent et Lausanne	plante entière
Arrayexpress <sup>5</sup>	CATMA	Little	Comparaison d'un échantillon d'ADNc d' <i>A. thaliana</i> avec lui-même	plante entière
Arrayexpress <sup>5</sup>	CATMA	Little	Comparaison des ADNc issus de 2 lots de plantes ayant poussé en parallèle	plante entière
Arrayexpress <sup>5</sup>	TIGR	White	Stress oxydatif	
GEO <sup>9</sup>	Agilent oligo	(Pufky <i>et al.</i> , 2003)	Cinétique de l'effet de l'IAA (auxine acide indole-3-acétique)	jeunes plantes
GEO <sup>6</sup>	Agilent oligo	(Pufky <i>et al.</i> , 2003)	Effets de six auxines différentes	jeunes plantes

<sup>4</sup> <http://www.plantcell.org/cgi/content/full/15/10/2320/DC1>

<sup>5</sup> [http://www.mpimp-golm.mpg.de/BR\\_reg\\_gene\\_expression/](http://www.mpimp-golm.mpg.de/BR_reg_gene_expression/)

<sup>6</sup> <http://www.plantphysiol.org/cgi/content/full/132/2/556/DC1>

<sup>7</sup> <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE911>

<sup>8</sup> ArrayExpress (Brazma *et al.*, 2003) : <http://www.ebi.ac.uk/arrayexpress/>

<sup>9</sup> GEO (Edgar *et al.*, 2002) : <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gds>

GEO <sup>6</sup>	ATH1	Alonso	Comparaison de la croissance en jours courts et en jours longs	apex
GEO <sup>6</sup>	ATH1	Alonso	Effet de l'éthylène sur des plantes étiolées de 3 jours	jeunes plantes
GEO <sup>6</sup>	ATH1	Nishimura	Expression des gènes du mutant <i>pmr4</i> (résistant au <i>mildew</i> ) dans des plants infectés et non infectés et au sein de plantes de phénotype sauvage	plante entière
GEO <sup>6</sup>	ATH1	Osborne	Analyse de feuilles, de tiges et de fleurs de plants d' <i>A. thaliana</i> poussés en chambre de culture ou en serre	feuilles, tiges, fleurs
GEO <sup>6</sup>	ATH1	Palatnik	Etude du mutant de développement des feuilles <i>jaw-D</i>	parties aériennes
GEO <sup>6</sup>	ATH1	Weigel	Cinétique de l'induction florale de différents allèles <i>leafy</i> , <i>flowering locus t</i> et constants dans les écotypes <i>Landsberg erecta</i> et <i>Colombia</i>	Apex
NASC <sup>10</sup>	ATH1	Barrena	Analyse de boutons floraux du mutant male stérile <i>ms1</i> à différents stades	boutons floraux
NASC <sup>7</sup>	ATH1	Boyce	Réponse au stress abiotique d'un mutant du gène <i>CDPK6</i>	jeunes plantes
NASC <sup>7</sup>	ATH1	Bramke	Analyse des mutants <i>sfr6</i> et <i>sfr2</i> sensibles au gel	parties aériennes
NASC <sup>7</sup>	ATH1	Brown	Cinétique de l'effet de l'auxine sur une suspension cellulaire	suspension cellulaire
NASC <sup>7</sup>	ATH1	Brueggemann	Réponse de l'écotype <i>Col-0</i> aux rayonnements UV Effet de l'inoculation de <i>P. parasitica</i> chez le mutant <i>rpp7</i> et le sauvage	parties aériennes
NASC <sup>7</sup>	ATH1	Buchanan-Wollaston	Analyse de trois mutants de réponse au stress et de deux écotypes sauvages <i>Col-0</i> et <i>Col5 glabrous</i> pendant la sénescence	feuilles
NASC <sup>7</sup>	ATH1	Campbell	Analyse du contrôle de la lignification par comparaison de plants d' <i>A. thaliana</i> sauvages et mutants	plante entière
NASC <sup>7</sup>	ATH1	Cornah	Analyse de deux mutants <i>knock-out</i> d'enzymes du cycle du glyoxylate : la malate synthase et l'isocitrate lyase	jeunes plantes
NASC <sup>7</sup>	ATH1	Deeken	Modification de l'expression des gènes en réponse au développement de tumeurs induites par <i>Agrobacterium tumefaciens</i> chez <i>A. thaliana</i> WS-2	tumeur
NASC <sup>7</sup>	ATH1	Evans	Effet de la température sur un mutant du métabolisme oxydatif	jeunes plantes
NASC <sup>7</sup>	ATH1	Filleur	Etude du rôle de <i>ANR1</i> dans l'induction de l'activité des méristèmes racinaires par le nitrate	racines
NASC <sup>7</sup>	ATH1	Greville	Comparaison de l'effet du saccharose sur le mutant <i>acn1-2</i> , incapable d'utiliser l'acétate comme source de carbone, et une plante sauvage	jeunes plantes
NASC <sup>7</sup>	ATH1	Hammond	Analyse du mutant <i>pho1</i> , déficient en phosphate, par rapport au sauvage	feuilles de rosette
NASC <sup>7</sup>	ATH1	Honys	Comparaison de l'expression dans les microspores haploïdes et le pollen mature	pollen
NASC <sup>7</sup>	ATH1	Jones	Effet de la mutation du gène <i>AtrbohC</i> sur l'expression des gènes dans les racines primaires	racines
NASC <sup>7</sup>	ATH1	Jordan	Analyse du mécanisme d'extinction de gènes par homologie ( <i>homology-dependent gene silencing</i> ) en comparant des mutants avec des plantes de phénotype sauvage	plante entière
NASC <sup>7</sup>	ATH1	Knight	Effet du froid sur les mutants sensibles au gel <i>sfr6</i> et <i>cls8</i>	jeunes plantes
NASC <sup>7</sup>	ATH1	Lloyd	Analyse du mutant <i>pho3</i>	feuilles
NASC <sup>7</sup>	ATH1	Millenaar	Induction de la croissance hyponastique par l'éthylène	pétiole
NASC <sup>7</sup>	ATH1	Okamoto	Effet de l'ABA (acide abscissique) sur la fermeture des stomates du mutant <i>gpa1</i> et des plantes de phénotype sauvage	plante entière
NASC <sup>7</sup>	ATH1	Pickett	Analyse du mutant sensible à la température <i>add3</i>	parties aériennes
NASC <sup>7</sup>	ATH1	Shirras	Comparaison des écotypes d' <i>A. thaliana</i> calcifuge <i>Col-4</i> et calcicole <i>Cal-0</i>	plante entière
NASC <sup>7</sup>	ATH1	Short	Effet du stress induit par l'ozone	plante entière
NASC <sup>7</sup>	ATH1	Smith	Cinétique du métabolisme du carbone en fonction du cycle jour/nuit	feuilles
NASC <sup>7</sup>	ATH1	De Torres Zabala	Cinétique de l'inoculation de différentes souches du pathogène <i>Pseudomonas syringae</i> sur <i>A. thaliana</i>	feuilles
NASC <sup>7</sup>	ATH1	Turner	Analyse du mutant de vascularisation <i>cov</i>	extrémités des tiges inflorescencielles
NASC <sup>7</sup>	ATH1	Villadsen	Etude de la voie de signalisation de l'hexokinase dans la détection du glucose par l'effet de différents sucres non métabolisables sur la croissance d' <i>A. thaliana</i> <i>Col-0</i>	jeunes plantes
NASC <sup>7</sup>	ATH1	Walters	Analyse d'un mutant du translocateur de triose-phosphate/phosphate de la membrane interne du chloroplaste	feuilles de rosette matures
NASC <sup>7</sup>	ATH1	Ward	Analyse de mutants hypersensibles et résistants à l'auxine pour caractériser des gènes impliqués dans l'activation des méristèmes des bourgeons axillaires	bourgeons axillaires
NASC <sup>7</sup>	ATH1	Warren	Effet du froid sur les trois mutants <i>sfr2</i> , <i>sfr3</i> et <i>sfr6</i>	parties aériennes
NASC <sup>7</sup>	ATH1	West	Effet d'un stress génotoxique à la bléomycine du mutant de réparation de l'ADN <i>atku80</i>	plante entière
NASC <sup>7</sup>	ATH1	Willats	Effet de l'herbicide isoxaben (2, 6-dichlorobenzonitrile) sur la biosynthèse de la pectine	suspension cellulaire
NASC <sup>7</sup>	ATH1	Yap	Effet de l'inoculation de <i>Gigaspora rosea</i> sur le développement des racines d' <i>A. thaliana</i>	racines

<sup>10</sup> NASC (Craigon *et al.*, 2004) : <http://affymetrix.arabidopsis.info/>

## B. L'étude de l'expression d'une famille de gènes

L'étude de l'expression des gènes d'une famille est soumise à des contraintes spécifiques provenant directement du taux important de ressemblance entre les séquences de par leur origine commune. Chacune des techniques d'analyse « à grande échelle » introduites précédemment peut être utilisée pour étudier l'expression des gènes d'une famille. Cependant, elles ne sont pas forcément appropriées.

### 1. Les méthodes d'analyse globale : étiquettes et AFLP

Par leur approche sans a priori, les techniques d'analyse en masse ne sont pas indiquées lorsqu'on étudie un groupe de gènes particulier telle qu'une famille. En effet, la part des transcrits des gènes ciblés représente au maximum 1 à 2% du transcriptome.

De plus, à cause de la proximité des séquences des gènes d'une famille, d'une part, la courte taille des étiquettes ne permet pas toujours de distinguer les différents transcrits provenant d'une même famille, et d'autre part, le mélange des transcrits dans les bandes observées en gel d'AFLP est possible, d'autant plus que des fragments issus de gènes de séquences proches risquent de migrer à la même vitesse.

Enfin, ces techniques ne sont pas appropriées pour l'étude des gènes faiblement exprimés.

### 2. Les puces à ADN

L'approche par *microarrays* peut être appliquée à l'étude de familles de gènes en déposant l'ensemble des gènes sur un filtre à haute densité par exemple. Cependant, la séquence des gènes d'une famille est caractérisée par un fort taux de ressemblance dépassant souvent 70% d'identité. Les risques d'artéfact sont donc réels. L'utilisation des *microarrays* pose le problème de l'hybridation croisée d'un transcrit sur les paralogues proches. Le biais est d'autant plus gênant si l'expression est très différente entre les paralogues. Par exemple, un gène 100 fois plus exprimé que son paralogue le plus proche pourra entraîner une sur-estimation de l'expression de ce dernier d'un facteur 10, si le taux d'hybridation croisée est d'environ 10%. A notre connaissance, une seule étude de famille de gènes a été menée par *microarrays* chez les plantes sur la famille des cytochromes P450 (Xu *et al.*, 2001). A travers cette étude, les auteurs montrent que des phénomènes d'hybridation croisée se produisent dès que le taux d'identité dépasse 70% entre deux paralogues. Et, si des zones de forte ressemblance locale existent pour des gènes partageant moins de 50% d'identité, l'expression est surévaluée. Le phénomène d'hybridation croisée peut être contourné avec l'utilisation de puces à oligonucléotides. Cependant, même la détermination d'oligonucléotides spécifiques

est difficile au sein d'une famille. Ainsi, il est nécessaire d'utiliser une autre technique pour étudier l'expression transcriptionnelle d'une famille de gènes.

### 3. La PCR quantitative en temps réel

La PCR quantitative en temps réel est une technique récente qui permet de mesurer l'expression des gènes via l'amplification spécifique du gène étudié par PCR. En utilisant la fluorescence comme système rapporteur, l'augmentation de la quantité d'ADN au cours de la PCR est enregistrée au fur et à mesure. Après la PCR, les quantités d'ADN sont comparées dans la partie exponentielle, moment pendant lequel l'augmentation de la quantité d'ADN est proportionnelle à la quantité initiale de matrice (Higuchi *et al.*, 1993; Heid *et al.*, 1996). Plusieurs revues font le point sur cette technique (Bustin, 2000; Bustin, 2002; Mingam *et al.*, 2003; Gachon *et al.*, 2004). Une présentation détaillée de cette technique est développée dans le chapitre suivant.

La technique de RT-PCR quantitative est utilisée pour mesurer l'expression transcriptionnelle des gènes, en remplacement ou en complément d'études par *Northern Blot* et par puces à ADN. De nombreuses études font appel à cette technique pour apporter de nouveaux éléments pour la compréhension de la régulation et la fonction des gènes. En biologie végétale, elles abordent différentes problématiques comme la signalisation cellulaire (Charrier *et al.*, 2002; Shimada *et al.*, 2003), la réponse aux stress biotiques (Berger *et al.*, 2002; Schenk *et al.*, 2003) et abiotiques (Charrier *et al.*, 2002; Kursteiner *et al.*, 2003), le fonctionnement du métabolisme (Yokoyama & Nishitani, 2001; Anterola *et al.*, 2002; Orsel *et al.*, 2002a; Suzuki *et al.*, 2002; Park *et al.*, 2003; Philippar *et al.*, 2003), l'édition des transcrits chloroplastiques (Peeters & Hanson, 2002), la réplication de l'ADN (Dambrauskas *et al.*, 2003) et la dégradation des protéines (Mladek *et al.*, 2003).

La PCR quantitative en temps réel est utilisée pour la caractérisation de l'expression transcriptionnelle d'ensembles de gènes appartenant à des familles multigéniques (Yokoyama & Nishitani, 2001; Charrier *et al.*, 2002; Orsel *et al.*, 2002a; Kursteiner *et al.*, 2003; Mladek *et al.*, 2003) ou non, par exemple les gènes appartenant à une même voie de biosynthèse (Anterola *et al.*, 2002), à une même voie de transduction (Shimada *et al.*, 2003) ou régulés en réponse à un même stimulus (Schenk *et al.*, 2003).

Par l'amplification spécifique d'une séquence cible grâce à la combinaison de deux oligonucléotides, la PCR quantitative en temps réel permet (1) d'étudier les gènes faiblement

exprimés et (2) de mesurer l'expression des gènes avec une meilleure spécificité que les techniques basées sur une seule hybridation. Ainsi, la PCR quantitative en temps réel nous a semblé l'approche la plus adaptée à l'étude de l'expression transcriptionnelle d'une famille de gènes. Le chapitre suivant présente les avantages et les limites de cette approche de la théorie à la pratique, avant de la valider pour l'appliquer à l'étude de l'expression de la famille des ARN hélicases à boîte DEAD d'*A. thaliana*.

**Résultats**

## La PCR quantitative en temps réel

Les deux articles de revue sur l'utilisation de la PCR quantitative en temps réel dans le domaine de la biologie végétale auxquels j'ai collaboré sont disponibles en annexe. Il s'agit des articles numérotés 1 et 2.

**Article 1: Mingam A, Gachon C and Charrier B.** Quantifying nucleic acids with fluorescent probes: contribution of real-time PCR to applied and fundamental plant research. *Advances in Plant Physiology*. Vol. 7, sous presse.

**Article 2: Gachon C, Mingam A and Charrier B.** Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany*. Sous presse.

### I. Introduction

Les premières tentatives visant à enchaîner des réactions de polymérisation successives d'ADN par catalyse enzymatique ou PCR (pour *Polymerisation Chain Reaction*) à des fins d'amplification datent des années 1985-1987 (Saiki *et al.*, 1985; Mullis *et al.*, 1986; Mullis & Faloona, 1987). A l'origine, la réaction était catalysée par le fragment Klenow de l'ADN polymérase I d'*E. coli*. Par la suite, le protocole original a été amélioré (Saiki *et al.*, 1988) en remplaçant cette ADN polymérase thermolabile par une ADN polymérase thermostable, la *Taq* polymérase, isolée de la bactérie thermophile *Thermus aquaticus* (Chien *et al.*, 1976). Cette innovation dans la manière d'amplifier *in vitro* un segment spécifique d'ADN a révolutionné la biologie moléculaire à un point tel que Kary B. Mullis a reçu pour cette invention le prix Nobel de chimie en 1993. En effet avant 1985, la seule façon d'obtenir un grand nombre de copies identiques d'un fragment d'ADN était de multiplier des bactéries transformées avec un plasmide comportant la séquence voulue, puis d'extraire et purifier les plasmides. Avec l'invention de la PCR, le temps nécessaire à la production de grandes quantités d'un fragment d'ADN spécifique a été réduit de plusieurs jours à quelques heures.

Le principe de la réaction de PCR est basé sur l'hybridation de courts fragments d'ADN simple brin sur une matrice d'ADN simple brin de séquence complémentaire. Ces oligonucléotides sont situés aux extrémités du fragment à amplifier. Ils servent d'amorces à une ADN polymérase ADN dépendante qui synthétise le brin complémentaire de la molécule d'ADN simple brin à laquelle ils sont hybridés. La matrice d'ADN est tout d'abord dénaturée par la chaleur en présence d'un large excès des deux oligonucléotides et des quatre dNTP. Le mélange réactionnel est refroidi jusqu'à une température permettant l'hybridation des amorces au niveau de leur séquence-cible. L'ADN polymérase allonge ensuite les amorces

par l'adjonction de nouveaux nucléotides complémentaires de la matrice d'ADN. Ce cycle de dénaturation, hybridation et synthèse d'ADN est ainsi répété plusieurs dizaines de fois. Comme les fragments d'ADN synthétisés au cours d'un cycle servent de matrice au cycle suivant, la quantité du fragment d'ADN ciblé s'accroît exponentiellement au cours de la PCR. Typiquement, le facteur d'amplification d'une séquence par une réaction de PCR est de l'ordre de  $10^6$  ( $2^n$  avec  $n = 20$ ). L'amplification est exponentielle pendant une grande partie de la PCR, puis la réaction se ralentit au cours des derniers cycles et atteint un plateau. La quantité finale de produit amplifié n'est alors plus proportionnelle à la quantité initiale de séquence-cible.

Simple, rapide et fiable, la PCR donne de très bons résultats qualitatifs. Ses applications sont nombreuses : clonage moléculaire, mutagenèse dirigée, criblage de banques d'ADN génomique et d'ADNc, analyse de l'expression transcriptionnelle des gènes mais aussi analyse moléculaire au niveau d'une cellule unique, diagnostic et suivi de maladies (cancers, infections et parasitoses) et détection de contaminants viraux, bactériens, fongiques et OGM dans les semences et les produits alimentaires.

Pour appliquer la PCR à l'étude de l'expression des gènes, une étape supplémentaire, précédant l'étape d'amplification, est nécessaire : il s'agit de la conversion des ARNm en ADN, dits ADN complémentaires ou ADNc. Cette étape fait appel à une autre ADN polymérase, la transcriptase inverse ou RT (*Reverse Transcriptase*), qui est ARN dépendante. Elle synthétise un brin d'ADN complémentaire d'une molécule d'ARN à partir d'une zone double brin créée par l'hybridation d'une amorce. La RT-PCR est donc à la base de l'étude de l'expression d'un gène par amplification moléculaire. Utilisée au départ pour déterminer l'expression ou l'absence d'expression d'un gène dans un échantillon donné, des techniques ont été développées pour permettre de quantifier l'expression des gènes par PCR. Pour l'adapter à ce but, il faut contourner le problème de la non-linéarité de l'amplification en fin de réaction. C'est pourquoi, pour comparer l'intensité des signaux entre deux réactions de PCR, il est nécessaire de procéder aux mesures au cours de la phase exponentielle.

Dans un premier temps, des études cinétiques ont été réalisées en prélevant dans le tube à intervalles réguliers des fractions aliquotes du mélange réactionnel. Après migration des fragments amplifiés à chaque prélèvement dans un gel coloré avec un agent fluorescent se liant à l'ADN (bromure d'éthidium ou SYBR-Green® par exemple), l'intensité des bandes est observée sous un éclairage excitant dans l'UV. La quantification peut également se faire par un marquage radioactif du produit ou par une détection immunologique. Ensuite la



représentation de l'intensité du signal en fonction du nombre de cycle de PCR permet de déterminer l'amplitude de la phase exponentielle de la réaction (entre 15 et 25 cycles de température en général). Les points des échantillons compris dans cet intervalle sont alors comparés aux points correspondants du standard externe. Cette technique est couramment utilisée, on y fait référence par le terme RT-PCR semi-quantitative (Ferre, 1992).

Des sophistications dans l'appréciation de la quantité initiale de séquence-cible ont été développées en introduisant un standard interne dans la réaction (Raeymaekers, 1993). Le standard de concentration connue, en compétition avec la cible, est ajouté dans le milieu réactionnel avant le début de la réaction de RT-PCR. Ce standard doit présenter des caractéristiques proche du produit à doser pour permettre son amplification avec la même efficacité que la séquence-cible. Plusieurs tubes contenant différentes dilutions de l'échantillon en présence de la même quantité de standard sont préparés en parallèle. Des prélèvements sont réalisés à intervalles réguliers pendant la PCR. Après migration des fragments amplifiés dans un gel, l'intensité des bandes est comparée. Lorsque l'intensité des bandes correspondant à la séquence-cible et au standard sont identiques, la quantification est réalisée. Cette méthode est appelée RT-PCR compétitive (Becker-Andre & Hahlbrock, 1989; Gilliland *et al.*, 1990). La RT-PCR non compétitive procède de la même façon, mais la quantification se déroule avant la phase de compétition des deux produits (Freeman *et al.*, 1999). Ces approches cinétiques sont intéressantes mais elles présentent des inconvénients techniques comme celui d'opérer des prélèvements successifs qui peuvent facilement être à l'origine de contamination entre des tubes voisins et de variation dans le déroulement de la PCR. En effet, selon le temps mis pour réaliser un prélèvement, une évaporation peut se produire, modifiant la concentration du milieu réactionnel et potentiellement l'efficacité de la réaction. Pour s'affranchir de ce problème, il est possible d'utiliser plusieurs tubes identiques et de les retirer de l'appareil au fur et à mesure. Cependant cette méthode a pour conséquence de multiplier les tubes donc le prix et la quantité nécessaire d'échantillon par analyse. C'est pourquoi, dès les années 1990, de nouvelles méthodes permettant l'enregistrement de l'amplification de la séquence-cible au cours de la PCR sans intervenir directement sur les tubes ont été mises au point.

En 1993, Higuchi et ses collaborateurs (Cie Roche) décrivent le premier dispositif expérimental d'enregistrement en temps réel de l'amplification d'une séquence par PCR (Higuchi *et al.*, 1993). Le dispositif consiste à éclairer simultanément les 96 puits du bloc chauffant d'un appareil PCR avec une lumière ultraviolette pendant toute la durée des cycles

thermiques. Un agent intercalant de l'ADN, le bromure d'éthidium (BrEt), est présent dans le milieu réactionnel. Sous l'action des UV, le BrEt lié à l'ADN double brin émet une fluorescence, captée et enregistrée par un système vidéo. Les images vidéo de la plaque sont enregistrées à chaque cycle pendant la phase d'élongation, stockées sur le disque dur d'un ordinateur puis analysées. Les auteurs de la publication montrent que la cinétique de l'augmentation de la fluorescence au cours du temps est directement liée au nombre initial de copies de la séquence-cible. Plus le nombre de cycles nécessaires pour obtenir un niveau de fluorescence détectable est faible, plus la séquence-cible est abondante au départ. Dès cette publication, les bases de la PCR quantitative en temps réel sont posées. Le protocole préconise l'amplification d'un court fragment d'ADN et un cycle à 2 températures. L'hybridation des amorces et l'élongation sont alors réalisées à la même température. Il est possible de détecter et quantifier de façon fiable et reproductible de 100 à  $10^9$  molécules d'ADN simple brin du gène cible parmi 300 ng d'ADN génomique humain, soit l'équivalent du contenu en ADN de 40.000 cellules (Higuchi *et al.*, 1993). Les auteurs insistent sur l'importance d'éviter la formation de dimères d'amorces pour garantir la validité de la mesure. Ils annoncent également la poursuite de leur recherche sur d'autres agents intercalants et fluorescents et sur des stratégies de détection alternatives utilisant la propriété de l'activité exonucléase 5'-3' de la *Taq* polymérase associée à des sondes fluorescentes.

Suite à cette publication, paraît en 1996 dans *Genome Research* un article intitulé « Real-time quantitative PCR » (Heid *et al.*, 1996). Il s'agit des résultats d'une collaboration entre deux entreprises américaines : Genentech et Applied Biosystems (division de Perkin Elmer Corporation). Ils présentent une nouvelle méthode permettant de mesurer l'accumulation du produit PCR via une sonde fluorogène doublement marquée : la sonde TaqMan. La fluorescence des deux marqueurs est inhibée par un phénomène d'extinction (*quenching*). Pendant l'amplification, l'ADN polymérase dégrade la sonde et ainsi libère les fluorophores qui émettent un rayonnement, enregistré par le système vidéo. Cette approche permet d'augmenter la spécificité du signal, en effet, la fluorescence émise provient uniquement du produit amplifié. De plus, il est possible de co-amplifier plusieurs produits repérés par des sondes marquées de fluorophores différents.

En 1997, c'est au tour de Roche de publier les résultats obtenus avec son appareil : le Light Cycler (Wittwer *et al.*, 1997). Il utilise un autre système de PCR : la PCR en capillaires de verre chauffés et refroidis par un flux d'air. Ce système existait déjà pour la PCR classique. Il

présente l'avantage de s'affranchir de l'inertie thermique des blocs chauffants et ainsi permettre l'amplification d'une séquence en une demi-heure environ au lieu de deux heures. Cet appareil enregistre également plusieurs longueurs d'ondes.

Actuellement, il existe au moins une dizaine de machines différant par le système d'émission et de détection de la fluorescence (lampe halogène, fibres optiques), le support de la réaction (tubes ou capillaires), le nombre d'échantillons analysable (32, 96 ou 384), la possibilité de réaliser ou non plusieurs réactions dans un même tube (multiplex), le nombre de longueurs d'onde mesurées, la possibilité d'effectuer en parallèle plusieurs programmes de PCR différents.

## II. Utilisations de la PCR quantitative en temps réel en biologie végétale

En plus de l'étude de l'expression transcriptionnelle d'un ou plusieurs gènes (Introduction IV.B.3.), la PCR quantitative en temps réel a trouvé de nombreuses applications en biologie végétale, couvrant divers aspects et thématiques de ce domaine.

### A. Sécurité alimentaire : détection d'OGM et de contaminants fongiques et bactériens

Pour sa sensibilité et ses propriétés quantitatives, la PCR en temps réel a été incluse dans les protocoles de contrôle-qualité utilisés couramment dans le domaine de la sécurité alimentaire, et plus particulièrement pour la détection de contaminants OGM, fongiques et bactériens dans la nourriture.

Elle a été utilisée avec succès pour la détection dans la nourriture et les stocks de semences d'organismes génétiquement modifiés issus de la transgénèse végétale, notamment plusieurs variétés de maïs et de soja (Vaitilingom *et al.*, 1999; Locatelli *et al.*, 2000; Hubner *et al.*, 2001; Ahmed, 2002 ; Brodmann *et al.*, 2002; Kuribara *et al.*, 2002; Terry *et al.*, 2002; Wiseman, 2002; Hernandez *et al.*, 2003). Le principe de ces tests est basé sur la détection et la mesure de l'abondance d'un fragment d'ADN caractéristique de la construction insérée dans le génome de la plante modifiée. La plus grande difficulté pour mettre en œuvre ces tests est d'avoir accès à cette séquence. En effet, toutes les variétés de semences OGM sont protégées par des brevets interdisant les expérimentations et donc le séquençage de la construction introduite. La source principale d'information reste donc le texte du brevet. Et, même quand ce texte est accessible, la description détaillée de la construction, comportant la séquence, est rarement précisée.

Les *Fusarium* sont des champignons attaquant les céréales. Ils secrètent des toxines pouvant persister dans la nourriture. Leur absence dans les stocks de semences est strictement suivie et contrôlée pour en assurer la qualité phytosanitaire. Des tests de détection de *Fusarium* dans les stocks de céréales par PCR classique et RT-PCR ont été d'abord développés (Niessen & Vogel, 1998; Doohan *et al.*, 1999), puis des tests utilisant la PCR quantitative en temps réel ont été mis en place (Edwards *et al.*, 2001; Schnerr *et al.*, 2001).

La détection de bactéries pathogènes dans les aliments est un test clé de la qualité sanitaire des produits de consommation destinés à la nutrition humaine et animale. Un test de détection de *Listeria* dans les choux, ingrédient entrant dans la composition du *coleslaw*, plat à l'origine d'une recrudescence des listérioses, a fait l'objet d'une étude en Angleterre (Hough *et al.*, 2002).

Pour toutes ces utilisations et notamment sur des échantillons provenant de grands volumes comme les stocks de semences, la difficulté principale dans l'application de ces tests est la représentativité de l'échantillon analysé, conduisant à l'acceptation ou au refus d'un lot devant le résultat obtenu sur le prélèvement. Ce problème d'échantillonnage n'est certes pas propre à la PCR quantitative en temps réel, mais il est exacerbé par la sensibilité de cette technique.

## B. Biologie Moléculaire

### 1. Gènes et transcrits végétaux utilisés comme standards

Pour assurer un meilleur suivi des expérimentations de PCR et RT-PCR quantitative en temps réel, l'utilisation de standards internes d'origine exogène peut être un choix judicieux. J'ai relevé des exemples de gènes et de transcrits végétaux utilisés dans ce but : l'ARN de la RuBisCo servant de contrôle exogène dans les expériences de RT-PCR quantitative sur des cellules animales et bactériennes (Smith *et al.*, 2003) et le gène codant l'acetyl-CoA carboxylase du Colza (*Brassica napus*) comme référence pour des expériences de détection d'OGM dans des échantillons de nourriture par PCR quantitative (Hernandez *et al.*, 2001).

### 2. Transgenèse

La PCR quantitative en temps réel a également été utilisée pour estimer le nombre d'insertions d'un transgène après transformation de plants de maïs et de tomate par *Agrobacterium tumefaciens* (Ingham *et al.*, 2001 ; Mason *et al.*, 2002). D'après les auteurs, cette méthode fournit d'aussi bons voire de meilleurs résultats qu'avec une technique classique par *Southern blot*.

### 3. Outil de validation

Une validation de la technique SAGE a été réalisée par PCR quantitative en temps réel (Ekman *et al.*, 2003). Il a été signalé l'importance de confirmer les résultats obtenus par *microarrays* par d'autres techniques, notamment la PCR quantitative en temps réel (Wurmbach *et al.*, 2003). Elle a notamment été utilisée en complément d'études d'expression transcriptionnelle par *microarrays* chez *A. thaliana* (Faccioli *et al.*, 2002; Maguire *et al.*, 2002; Hammond *et al.*, 2003; Wang *et al.*, 2003a)

### C. Phytopathologie

De nombreuses interactions plante-microorganisme font l'objet d'études utilisant la PCR quantitative en temps réel. Certaines sont focalisées sur les mécanismes d'interaction (Skaf *et al.*, 2000; Cooper, 2001; Guescini *et al.*, 2003; McMaugh & Lyon, 2003; Morissette *et al.*, 2003; Schenk *et al.*, 2003). D'autres sont plutôt orientées vers la détection et l'estimation de l'ampleur de l'infection (Weller *et al.*, 2000; Postma *et al.*, 2001; Schnerr *et al.*, 2001) ou de la diversité fongique (Heuser & Zimmer, 2002; Stubner, 2002).

### III. Principe et définitions

La richesse des approches impliquant la PCR quantitative en temps réel, la diversité des modèles de machine utilisés, les différents systèmes de fluorescence possibles (agent intercalant ou sondes oligonucléotidiques) et le choix de réaliser la réaction de RT-PCR en un seul tube ou en deux tubes séparés sont autant de sources de variation des protocoles mis en œuvre (Bustin, 2000; Bustin, 2002; Klein, 2002).

Dans ce chapitre, je décris le système expérimental présent au laboratoire, les principes généraux de la quantification par PCR quantitative en temps réel, valables quels que soient l'appareil et la méthode utilisés, et mes choix expérimentaux pour le système d'émission de la fluorescence, la gamme-étalon, l'unité de mesure et la conception des amorces.

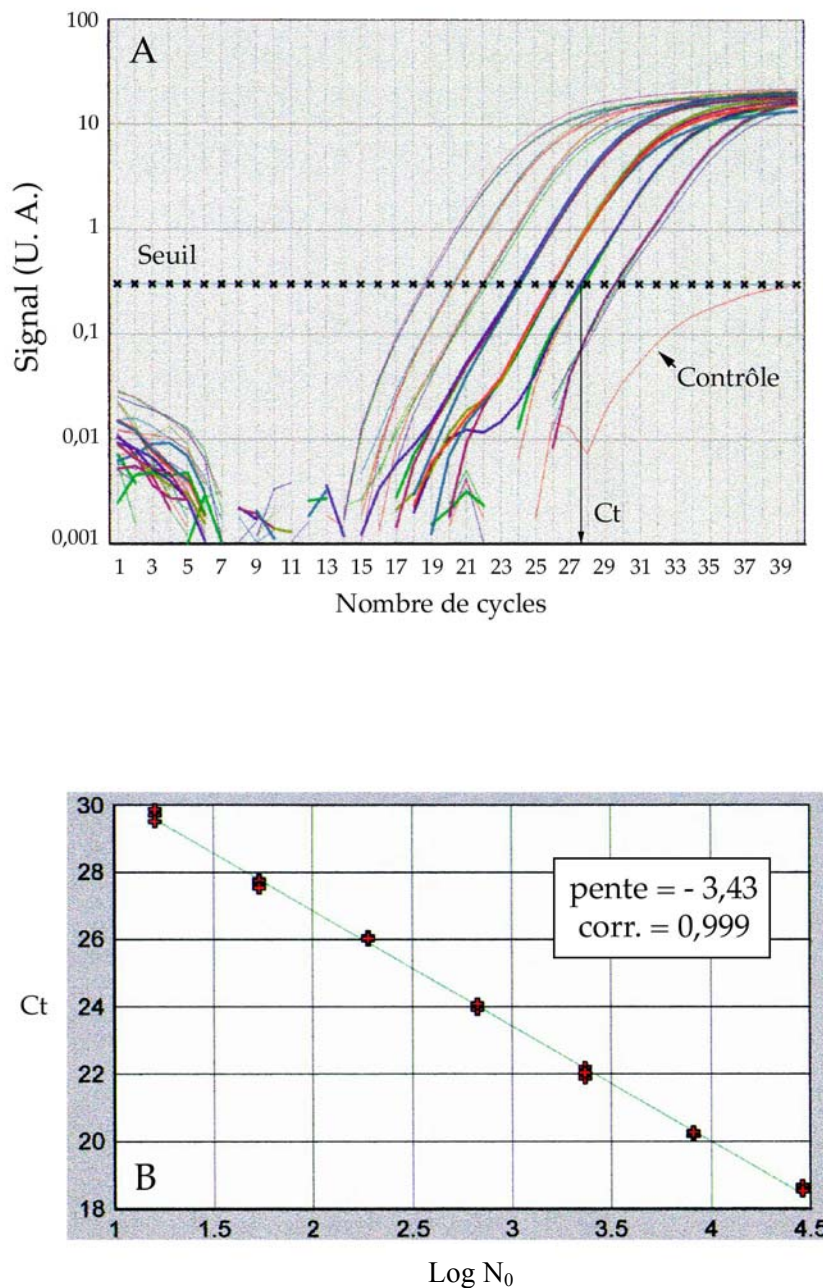
#### A. Dispositif expérimental

Le dispositif expérimental disponible au laboratoire est un GeneAmp® 5700 Sequence Detection System (Applied Biosystems). Il est composé d'un thermocycleur classique associé à un système d'enregistrement de la fluorescence, interfacé sur un ordinateur PC qui contrôle le fonctionnement de l'ensemble.

Une lampe halogène produit, grâce à un filtre, une lumière monochromatique centrée à 485 nm qui est transmise optiquement jusqu'à la plaque de PCR. Le SYBR Green®, fluorophore présent dans le mélange réactionnel, est excité à cette longueur d'onde et émet une fluorescence verte à 520 nm. Le rayonnement émis est capté par un système de détection vidéo et les images de la plaque sont stockées dans la mémoire de l'ordinateur.

Le SYBR-Green® est une molécule qui se fixe dans le petit sillon de l'hélice de l'ADN. Son rendement quantique de fluorescence (intensité du rayonnement émis) augmente significativement lorsqu'il est fixé à de l'ADN double-brin. Ainsi, grâce à cette molécule et au système d'enregistrement de la fluorescence, la progression de l'amplification est suivie à chaque cycle et en temps réel. Un puits contrôle ne contenant pas d'ADN permet de visualiser la fluorescence propre du SYBR-Green® non fixé.

Le logiciel Sequence Detection System (Applied Biosystems) permet d'une part de programmer et de contrôler le déroulement des expériences de PCR quantitative et d'autre part, d'en stocker et exploiter les données. L'interface graphique est divisée en trois parties : la première pour saisir les données concernant les puits (position, couple d'amorces, type d'échantillon), la seconde pour saisir les conditions de PCR et la troisième pour afficher l'analyse des résultats.



**Figure 2** Courbes d'amplification du gène *AtRH24* à partir de différentes quantités d'ADN génomique d'*A.thaliana* et gamme-étalon correspondante.

A. La fluorescence des puits au cours de la PCR est représentée en fonction du nombre de cycles. Chaque faisceau regroupe 4 courbes correspondant à 4 amplifications réalisées avec la même quantité initiale d'ADN. Le seuil de fluorescence, symbolisé par une ligne horizontale, permet de déterminer le cycle-seuil (Ct) de chaque courbe. La fluorescence du contrôle sans ADN, indiquée par une flèche, est inférieure au seuil retenu pour l'analyse.

B. Gamme-étalon construite en exprimant le Ct de chaque amplification en fonction du logarithme de la quantité initiale d'ADN génomique déposée dans chaque puits ( $\text{Log } N_0$ ). Le coefficient de corrélation (corr.) et la pente de la droite sont indiqués sur le graphique. D'après la valeur de la pente, l'efficacité de cette PCR est de 95,7%.



## B. Quantification de l'ADN par la détection de fluorescence

### 1. Cycle-seuil et nombre de copies initial

A la fin de la PCR, le logiciel représente graphiquement l'augmentation de la fluorescence de chaque puits au cours des cycles successifs (Figure 2A). Le nombre de cycles étant placé en abscisse et le logarithme de l'intensité de la fluorescence en ordonnée, les amplifications suivent une courbe sigmoïde en 3 phases. Pendant la première phase (pouvant durer 15 à 25 cycles), la fluorescence du puits contenant de l'ADN ne se démarque pas du bruit de fond. Puis l'augmentation de la fluorescence passe par une phase exponentielle (deuxième phase, 5 cycles environ) avant de ralentir et atteindre un plateau (dernière phase).

Le cycle-seuil ou  $C_t$  (*threshold cycle*) correspond au cycle au cours duquel la fluorescence d'un échantillon devient significativement différente du bruit de fond. Ce seuil, déterminé par l'utilisateur après analyse des profils de fluorescence, permet de comparer toutes les amplifications pendant la phase exponentielle. Si toutes les conditions de PCR préconisées par Applied Biosystems sont respectées, il existe une relation linéaire entre le nombre de copies d'un gène présent initialement dans le puits et le cycle-seuil.

Si  $N$  est le nombre de copies du gène au cycle  $C_n$  de la PCR et  $N_0$  le nombre de copies présentes au départ dans le tube, le nombre de copies du gène au cycle  $C_n$  peut être exprimé par la formule :

$$N = N_0 \cdot (1 + E)^{C_n} \quad (1)$$

où  $E$  est l'efficacité de la PCR (avec  $E \leq 1$ ).

Le nombre de molécules accumulées au cycle-seuil  $C_t$  quelque soit  $N_0$  est donc :

$$N = N_0 \cdot (1 + E)^{C_t} \quad (2)$$

Au cycle-seuil, l'intensité de la fluorescence est la même pour tous les puits. Donc, au cycle-seuil, le nombre de copies est identique quelle que soit la quantité initiale de matrice. Ainsi, pour deux nombres de molécules initiaux différents,  $N_1$  et  $N_2$ , on peut écrire :

$$N_1 \cdot (1 + E)^{C_{t1}} = N_2 \cdot (1 + E)^{C_{t2}}$$

soit :

$$\frac{N_1}{N_2} = \frac{(1 + E)^{C_{t2}}}{(1 + E)^{C_{t1}}}$$

ou encore :

$$\frac{N_1}{N_2} = (1 + E)^{(C_{t2} - C_{t1})} \quad (3)$$

La différence de cycle-seuil ( $Ct_2 - Ct_1$ ) est donc fonction du rapport des nombres de molécules initiaux respectifs, et, pour  $E = 1$ , si  $Ct_2 - Ct_1 = 1$ , alors  $N_1/N_2 = 2$ .

C'est-à-dire que s'il existe une différence d'un cycle entre les  $Ct$  de deux puits (par exemple si  $Ct_1 = 24$ ,  $Ct_2 = 25$ ), on peut en déduire que le puits n°1 contenait au départ deux fois plus de copies du gène amplifié que le puits n°2.

Ainsi, en comparant la valeur du cycle-seuil obtenu pour un gène donné dans un échantillon aux valeurs de la gamme-étalon contenant des quantités connues de ce gène, il est possible de déterminer le nombre de copies du gène présent initialement dans cet échantillon.

## 2. Equation de la gamme-étalon

Pour mesurer précisément le nombre de copies initial d'un échantillon ( $N_0$ ), il est plus facile d'utiliser une représentation linéaire reliant le  $Ct$  à  $N_0$ .

Le nombre de molécules accumulées au cycle-seuil  $Ct$  est exprimée par l'équation (2) :

$$N = N_0 \cdot (1+E)^{Ct}$$

D'où on en déduit :

$$\log N = \log N_0 + Ct \cdot \log (1+E) \quad (4)$$

En exprimant le cycle-seuil ( $Ct$ ) en fonction du logarithme de la quantité de matrice initiale ( $\log N_0$ ), on obtient la droite d'équation suivante :

$$Ct = -\frac{1}{\log(1+E)} \cdot \log N_0 + \frac{\log N}{\log(1+E)} \quad (5)$$

Cette droite est obtenue à partir des valeurs de  $Ct$  des quantités standards (Figure 2B). Elle sert de gamme-étalon pour réaliser les mesures des échantillons.

## 3. Relation pente-efficacité de la PCR

Lors des mesures, il est très important de se placer à une efficacité de PCR maximale (*i.e.*  $E = 1$ ). En effet, si  $E$  est différent de 1, la relation entre le  $Ct$  et la quantité initiale  $N_x$  est modifiée, et cela fausse l'estimation de celle-ci. Avant de procéder à la quantification des échantillons, il est donc indispensable de vérifier l'efficacité de la PCR.

D'après l'équation (5), la pente de la droite correspond à l'expression :

$$pente = -\frac{1}{\log(1+E)} \quad (6)$$

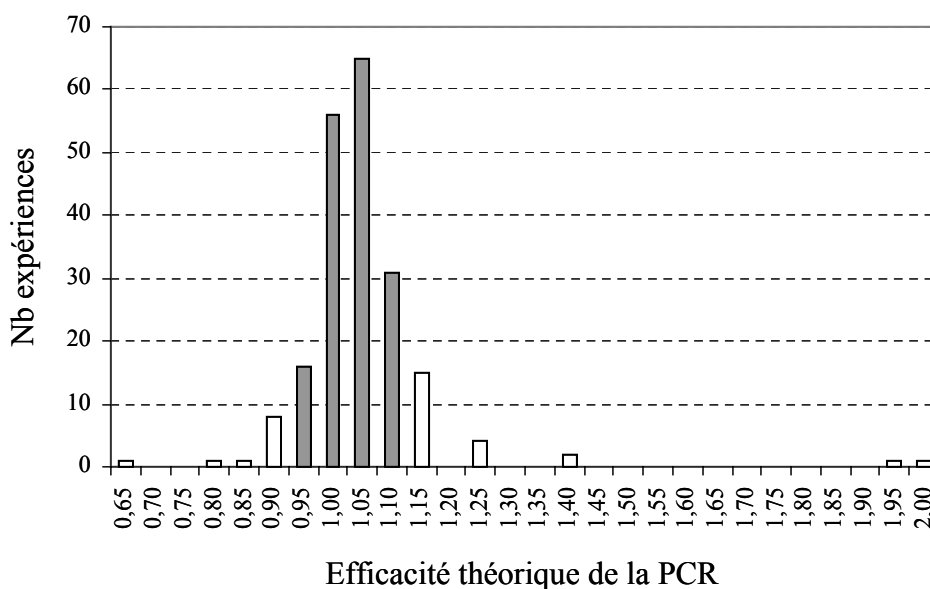
Donc, lorsque l'efficacité de la PCR est de 100% ( $E = 1$ ), la pente est théoriquement égale à  $-\frac{1}{\log 2}$ , c'est-à-dire à -3,32.

La pente de la gamme-étalon permet de déduire l'efficacité de la PCR selon l'équation suivante :

$$E = 10^{-1/pente} - 1 \quad (7)$$

#### 4. Efficacité des amorces et variations expérimentales de la gamme étalon

Mathématiquement, pour que  $E$  ne soit pas supérieure à 1, la pente de la droite doit être inférieure à -3,32. Cependant, comme la pente est calculée à partir de valeurs expérimentales, elle fluctue autour de la valeur -3,32 dans les deux sens. Elle peut donc être égale à -3,30 comme à -3,34. Nous avons accepté une variation expérimentale de l'efficacité de la PCR de plus ou moins 10%, c'est-à-dire les pentes comprises entre -3,60 et -3,10. Sur les 192 gammes réalisées avec des amorces spécifiques, 172 expériences (soit 90%) présentaient une pente appartenant à cet intervalle (Figure 3).



**Figure 3 Répartition des expériences de PCR quantitative selon leur efficacité théorique.**

L'efficacité de la PCR est déduite de la pente de la gamme-étalon d'après l'équation (6). Ce graphique représente les résultats obtenus avec l'ensemble des couples d'amorces testés ( $n = 202$ ). Il comprend les valeurs obtenues avec des couples d'amorces spécifiques et avec des couples éliminés pour non-spécificité ou mauvaise efficacité. La valeur indiquée pour l'efficacité correspond à la limite supérieure de chaque classe. Ainsi, l'indication « 0,65 » équivaut à l'intervalle ]0,6 ; 0,65]. Les histogrammes gris correspondant aux expériences retenues pour l'analyse des résultats.

## C. Choix expérimentaux

### 1. Le système d'émission de fluorescence

Différents systèmes de détection sont disponibles sur le marché. Le plus simple, utilisé assez fréquemment en biologie végétale, est basé sur le SYBR Green® qui émet une fluorescence lorsqu'il est lié à l'ADN double brin. Ce système est indépendant de la séquence amplifiée. Si des amplifications non spécifiques se produisent, elles sont également détectées. Pour pallier cet inconvénient, d'autres systèmes ont été développés. Ils font appels au phénomène d'extinction de la fluorescence (*quenching*) ou d'excitation (*Fluorescence Resonance Energy Transfer*) par proximité physique de deux fluorophores. Dans le cas du *quenching*, les deux molécules sont séparées au cours de l'amplification pour permettre l'émission de la fluorescence. Cette séparation peut être réalisée sous l'action de l'ADN polymérase au cours de l'amplification par la dégradation de la sonde nucléotidique doublement marquée (TaqMan®). Lorsque les fluorophores sont maintenus à proximité l'un de l'autre par une structure en épingle à cheveux, c'est l'hybridation de la sonde sur sa séquence cible qui sépare les deux molécules (Molecular Beacons®, Scorpion®). Dans le cas du FRET, le transfert d'énergie est réalisé lors de l'hybridation par la juxtaposition de deux oligonucléotides (Hybridization Probe).

Les systèmes de sonde se valent en termes de spécificité et de sensibilité. Cependant, le SYBR Green® donne un signal plus intense que les sondes. En effet, plusieurs molécules de SYBR Green® s'intercalent dans chaque fragment amplifié, alors que le marquage par les sondes est un marquage unimoléculaire. Le SYBR Green® permet donc une plus grande sensibilité des mesures. D'autre part, l'usage de sondes spécifiques pour chaque gène est plus coûteux dans le cadre de l'étude de l'expression d'une famille multigénique que le SYBR Green®, marqueur fluorescent « universel ».

En prenant soin de définir des amorces spécifiques, le SYBR Green® est plus performant que les sondes, pour un moindre coût. C'est pourquoi j'ai retenu ce système pour mesurer l'expression transcriptionnelle de la famille des *AtRH*.

### 2. Gamme-étalon

La quantification de l'expression des gènes repose sur la gamme-étalon. Cette gamme est spécifique du couple d'amorces utilisé, c'est-à-dire du gène étudié. Lorsqu'on travaille sur un nombre restreint de gènes, il peut être avantageux d'utiliser une gamme d'ADNc pour chaque gène, comme c'est souvent le cas en recherche clinique. Cependant, l'utilisation des ADNc implique non seulement de disposer de l'ADNc cloné de l'ensemble des gènes étudiés

**Tableau 4 Valeurs moyennes des Ct de la gamme-étalon pour chaque couple d'amorces.**

La gamme de quantités initiales d'ADN est exprimée en équivalents-génome. La pente moyenne de la gamme-étalon est indiquée dans la deuxième colonne. Les résultats sont présentés sous la forme : Ct moyen  $\pm$  écart-type (nombre de valeurs). Les lignes grisées indiquent des gènes pour lesquels les Ct moyens sont inférieurs à ceux des autres gènes.

Gène	Pente	Quantité initiale d'ADN (eg)				
		740	2592	9070	31746	111111
<i>AtRH01</i>	-3,29	28,6 $\pm$ 2,1 (21)	26,8 $\pm$ 2,1 (21)	25,0 $\pm$ 2,1 (21)	23,3 $\pm$ 2,2 (21)	21,2 $\pm$ 2,0 (19)
<i>AtRH02</i>	-3,31	28,0 $\pm$ 1,7 (32)	26,3 $\pm$ 1,7 (32)	24,5 $\pm$ 1,7 (32)	22,8 $\pm$ 1,7 (28)	21,2 $\pm$ 1,3 (28)
<i>AtRH04</i>	-3,32	27,8 $\pm$ 1,2 (98)	26,0 $\pm$ 1,1(101)	24,2 $\pm$ 1,1 (97)	22,3 $\pm$ 1,0 (96)	21,0 $\pm$ 0,8 (67)
<i>AtRH05</i>	-3,23	28,2 $\pm$ 2,4 (20)	26,4 $\pm$ 2,2 (20)	24,6 $\pm$ 2,3 (20)	22,9 $\pm$ 2,4 (20)	21,1 $\pm$ 2,1 (19)
<i>AtRH06</i>	-3,36	28,0 $\pm$ 2,2 (25)	25,9 $\pm$ 2,1 (28)	24,1 $\pm$ 2,1 (28)	22,2 $\pm$ 2,0 (28)	21,3 $\pm$ 2,0 (20)
<i>AtRH08</i>	-3,22	28,1 $\pm$ 2,3 (15)	26,2 $\pm$ 2,3 (15)	24,5 $\pm$ 2,1 (15)	22,7 $\pm$ 2,0 (15)	21,9 $\pm$ 2,3 (9)
<i>AtRH12</i>	-3,27	28,1 $\pm$ 2,3 (19)	26,2 $\pm$ 2,2 (19)	24,4 $\pm$ 2,1 (19)	22,7 $\pm$ 2,2 (19)	21,4 $\pm$ 1,9 (11)
<i>AtRH14</i>	-3,45	30,0 $\pm$ 1,6 (6)	27,9 $\pm$ 1,5 (6)	26,0 $\pm$ 1,7 (6)	24,2 $\pm$ 1,4 (6)	22,7 $\pm$ 1,3 (6)
<i>AtRH15</i>	-3,20	26,3 $\pm$ 0,7 (8)	24,6 $\pm$ 0,5 (8)	22,8 $\pm$ 0,6 (8)	21,6 $\pm$ 0,7 (3)	
<i>AtRH19</i>	-3,29	29,2 $\pm$ 1,6 (21)	27,7 $\pm$ 1,8 (23)	25,8 $\pm$ 1,9 (23)	24,1 $\pm$ 1,8 (23)	22,2 $\pm$ 1,8 (16)
<i>AtRH20</i>	-3,19	29,1 $\pm$ 0,0 (4)	27,1 $\pm$ 0,1 (4)	25,4 $\pm$ 0,1 (4)	23,7 $\pm$ 0,2 (4)	22,1 $\pm$ 0,3 (4)
<i>AtRH21</i>	-3,25	28,9 $\pm$ 0,2 (4)	27,2 $\pm$ 0,1 (4)	25,3 $\pm$ 0,2 (4)	23,4 $\pm$ 0,3 (4)	21,9 $\pm$ 0,2 (3)
<i>AtRH23</i>	-3,26	29,1 $\pm$ 1,8 (27)	27,3 $\pm$ 1,7 (27)	25,4 $\pm$ 1,8 (28)	23,7 $\pm$ 1,8 (28)	21,8 $\pm$ 1,9 (24)
<i>AtRH24</i>	-3,21	29,6 $\pm$ 2,2 (33)	27,6 $\pm$ 2,3 (36)	25,8 $\pm$ 2,4 (34)	23,5 $\pm$ 2,3 (29)	22,3 $\pm$ 2,2 (26)
<i>AtRH30</i>	-3,50	30,1 $\pm$ 3,4 (10)	28,4 $\pm$ 2,9 (10)	26,6 $\pm$ 2,7 (10)	24,8 $\pm$ 2,5 (10)	23,0 $\pm$ 2,0 (10)
<i>AtRH34</i>	-3,32	28,6 $\pm$ 1,2 (12)	27,0 $\pm$ 1,1 (12)	25,2 $\pm$ 1,0 (12)	23,5 $\pm$ 1,0 (11)	23,4 $\pm$ 0,0 (1)
<i>AtRH38</i>	-3,59	31,6 $\pm$ 3,0 (16)	29,6 $\pm$ 2,7 (15)	27,1 $\pm$ 2,5 (16)	25,7 $\pm$ 2,8 (16)	24,9 $\pm$ 2,9 (12)
<i>AtRH40</i>	-3,44	29,3 $\pm$ 0,4 (4)	27,3 $\pm$ 0,3 (4)	25,4 $\pm$ 0,3 (4)	23,6 $\pm$ 0,2 (4)	21,8 $\pm$ 0,6 (4)
<i>AtRH42</i>	-3,26	30,0 $\pm$ 0,7 (4)	28,3 $\pm$ 0,1 (4)	26,3 $\pm$ 0,3 (4)	24,6 $\pm$ 0,2 (4)	23,0 $\pm$ 0,6 (4)
<i>AtRH56</i>	-3,35	30,1 $\pm$ 0,4 (8)	28,1 $\pm$ 0,5 (8)	26,3 $\pm$ 0,4 (8)	24,6 $\pm$ 0,5 (8)	23,1 $\pm$ 0,1 (3)
<i>AtRH58</i>	-3,23	27,2 $\pm$ 0,6 (8)	25,4 $\pm$ 0,6 (8)	23,6 $\pm$ 0,5 (8)	21,9 $\pm$ 0,5 (8)	20,9 $\pm$ 0,4 (2)
<i>ACT</i>	-3,04	25,2 $\pm$ 0,2 (9)	23,3 $\pm$ 0,2 (9)	22,7 $\pm$ 2,0 (11)	20,9 $\pm$ 2,0 (11)	19,8 $\pm$ 2,2 (8)
<i>EF1<math>\alpha</math>4</i>	-3,21	26,6 $\pm$ 2,5 (17)	24,7 $\pm$ 2,4 (17)	22,9 $\pm$ 2,4 (17)	21,3 $\pm$ 2,3 (17)	19,7 $\pm$ 2,3 (17)
<i>iAtSK</i>	-3,39	28,0 $\pm$ 0,1 (2)	26,1 $\pm$ 0,0 (2)	24,6 $\pm$ 0,2 (2)	22,9 $\pm$ 0,3 (2)	20,5 $\pm$ 0,1 (2)
<i>SUP</i>	-3,69	31,7 $\pm$ 0,2 (2)	29,6 $\pm$ 0,0 (2)	27,9 $\pm$ 0,1 (2)	24,9 $\pm$ 0,2 (2)	

mais surtout de devoir comparer des résultats basés sur des gammes construites à partir de matrices différentes. Pour cette raison, il nous a semblé plus judicieux de choisir l'ADN génomique pour construire la gamme-étalon. Cela nous permet de garder la même matrice quel que soit le gène étudié et d'avoir ainsi une référence homogène pour l'ensemble de nos résultats. Cependant, il est alors nécessaire d'éviter les introns et les bordures d'exons pour placer les amorces.

Les Ct observés pour une quantité d'ADN donnée sont similaires d'un couple d'amorces à l'autre, aux variations d'efficacité près (Tableau 4). Cette homogénéité confirme l'intérêt d'utiliser de l'ADN génomique pour réaliser la gamme-étalon. Une gamme caractéristique de l'ADN génomique d'*A. thaliana* pourrait être définie et utilisée comme référence pour toutes les expériences. Cependant, nous avons préféré utiliser les valeurs expérimentales de la gamme associée à chaque expérience de PCR quantitative pour analyser les résultats.

Trois couples d'amorces se démarquent de cette gamme « universelle » par leurs Ct plus faibles : il s'agit des amorces utilisées pour les gènes *ACT2/8*, *EF1 $\alpha$ -4* et *AtRH15* (lignes grisées du tableau ci-contre). Des Ct plus faibles indiquent que le nombre initial de copies est plus élevé, c'est-à-dire l'amplification simultanée de plusieurs gènes. C'est le cas pour l'actine, puisque ses amorces ont été conçues pour amplifier à la fois les gènes *ACT2* et *ACT8*. Ce résultat est donc cohérent. Le gène *EF1 $\alpha$ -4* appartient à une famille multigénique de 4 gènes (Axelos *et al.*, 1989). Il est possible que les amorces amplifient un ou plusieurs des autres membres de cette famille. Le gène *AtRH15* partage plus de 95% d'identité avec le gène *AtRH56*. Malgré les précautions prises pour s'assurer de la spécificité de ce couple d'amorces (cf même chapitre, IV.B. Spécificité des amorces), la valeur des Ct semble indiquer l'amplification des deux paralogues. D'ailleurs, son expression mesurée par PCR quantitative est largement supérieure à la valeur à laquelle on s'attend d'après le nombre d'EST et l'abondance relative des MPSS disponibles pour ce gène (cf chapitre Résultats, I.B.3. *Northern in silico*). Si l'on soustrait l'expression totale d'*AtRH56*, gène identique à 95%, de celle d'*AtRH15*, on retrouve une expression plus proche de la tendance globale donnée par les EST et les MPSS. L'expression transcriptionnelle mesurée pour *AtRH15* serait potentiellement contaminée, même si tous les contrôles effectués sur l'amplification par PCR : une seule bande sur gel, un seul pic en dissociation, une seule séquence, une efficacité de 100% et le test de spécificité avec les oligos croisés (Tableau 7), confirment une

amplification spécifique. Pour cette raison, nous avons retiré l'expression du gène *AtRH15* dans la suite des analyses.

### 3. Interprétation des résultats en nombre de copies

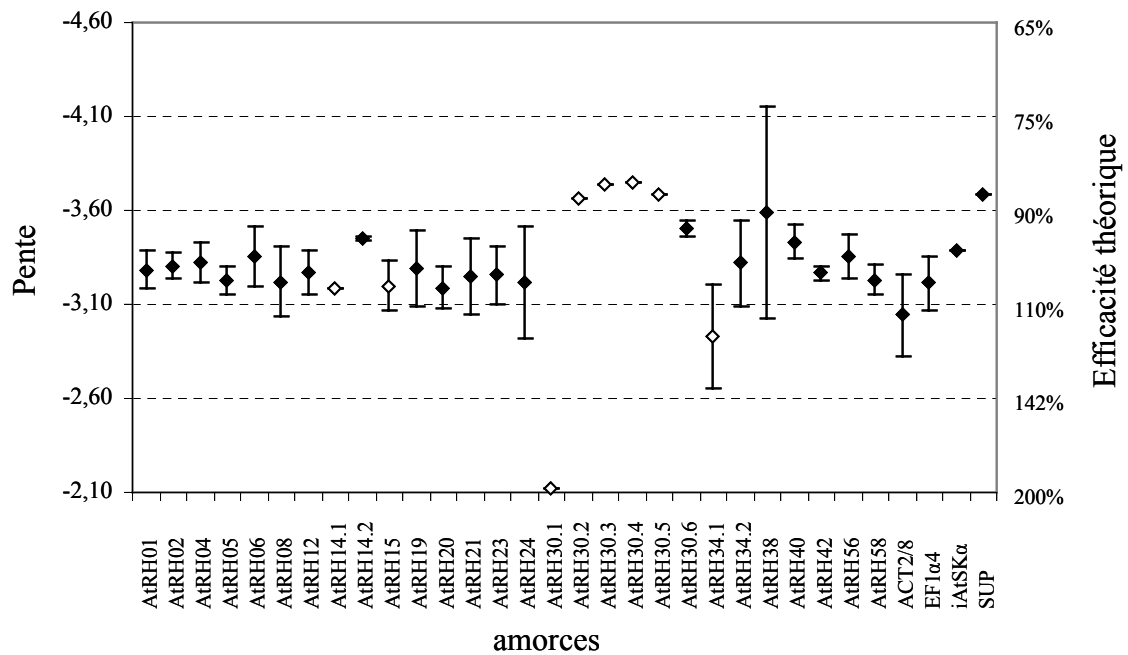
Le logiciel Sequence Detection System détermine les cycles-seuil (Ct) et les quantités initiales de matrice exprimées dans n'importe quelle unité. Cela peut être en unités arbitraires, en nanogrammes ou en nombre de copies de plasmide par exemple. Une étude précise en cytométrie de flux, confirmée par les données du séquençage, a permis d'évaluer la masse du génome diploïde d'*A. thaliana* à  $0,30 \pm 0,012$  pg (Arumuganathan & Earle, 1991). Il est donc possible d'exprimer en « équivalents génome » (eg), les quantités d'ADN génomique déposées pour la gamme en les divisant par la masse du génome haploïde d'*A. thaliana* (0,15 pg) ce qui signifie qu'un « équivalent génome » représente approximativement une copie du gène étudié. Cependant, comme au cours de nos expériences, la fluorescence d'une amplification réalisée sur de l'ADN génomique double brin est comparée à celle obtenue sur des ADNc simple brin, 1 équivalent-génome correspond à 2 copies d'ADNc simple brin. L'abondance des transcrits a été rapportée à la quantité d'ARN totaux utilisée à chaque RT (5 µg) afin de ramener les résultats à une teneur en ARN totaux similaire. Les résultats de PCR quantitative en temps réel sont donc exprimés en équivalents-génome par nanogrammes d'ARN totaux (eg.ng<sup>-1</sup> ARN totaux).

Pour chaque mesure, une gamme de dilutions de l'échantillon étudié est réalisée afin de s'assurer de la linéarité des résultats, c'est-à-dire qu'une dilution au 1/5<sup>e</sup> de la quantité d'ADN déposé doit se traduire par un nombre d'équivalents-génome cinq fois plus faible.

De plus, il est important de noter que, du fait de la nature exponentielle de la réaction de PCR, une différence d'un facteur 3 à un niveau d'expression faible (par exemple, entre 20 et 60 eg.ng<sup>-1</sup> ARN totaux) est aussi significative qu'à un fort niveau d'expression (entre 2000 et 6000 eg.ng<sup>-1</sup> ARN totaux, par exemple).

### 4. Choix des amorces

Une condition primordiale de la fiabilité des résultats en PCR quantitative est l'utilisation d'un couple d'amorces efficace à 100%. Un tel couple d'amorces est un couple dont le comportement permet une amplification strictement exponentielle de la séquence d'ADN cible, au moins pendant les premières phases de la PCR. Pour cela, plusieurs critères ont guidé la sélection des amorces.



**Figure 4 Variations expérimentales de la pente pour chaque couple d'amorces.**

Le graphique représente la pente moyenne des gammes-étalon obtenue pour chaque couple d'amorces. La barre d'erreur correspond à l'écart-type de chaque moyenne. Lorsqu'il n'y avait qu'une seule pente pour un couple donné, l'écart-type n'a pas été calculé. Les losanges blancs indiquent les couples d'amorces éliminés soit pour non spécificité (AtRH14.1 et AtRH15), soit pour mauvaise efficacité (AtRH30.1-5 et AtRH34.1).



Tout d'abord la probabilité de formation de structures secondaires et de dimères a été limitée pour éviter la perte de rendement de la PCR due au repliement et/ou à la dimérisation des amorces, particulièrement entre les extrémités 3'.

Les oligonucléotides doivent présenter également une  $T_m$  de  $59 \pm 1^\circ\text{C}$  d'après le calcul du logiciel Primer Express® (méthode du plus proche voisin) et générer un fragment de 50 à 200 pb maximum pour assurer une amplification efficace dans les conditions de la PCR quantitative.

Les couples retenus ne doivent pas pouvoir générer plusieurs amplifiats. En effet, la compétition de fixation des oligonucléotides sur différents sites perturbe nettement l'équilibre de la réaction, qui n'est alors plus représentative de la population du transcrit étudié. Leur spécificité vis-à-vis des gènes étudiés est testée virtuellement par rapport à toutes les séquences connues des ADNc d'*AtRH*.

Enfin, pour éviter une différence d'efficacité de PCR entre la gamme et les échantillons due à leur nature différente (ADN génomique ou ADNc), les amorces sont placées dans un même exon.

Chaque couple d'amorces est ensuite testé expérimentalement sur une gamme d'ADN génomique et sur un échantillon d'ADNc. L'efficacité moyenne de chaque couple d'amorces testé a été calculée sur l'ensemble des expériences de PCR quantitative (Figure 4). Plusieurs couples ne répondant pas aux critères d'efficacité (*AtRH30.1-5* et *AtRH34.1*) et de spécificité (*AtRH14.1* et *AtRH15*) ont été éliminés après ce test. Mis à part les amorces spécifiques du gène *AtRH38* qui présentent une forte variation (2 pentes sur 8 inférieures à -4,10), l'ensemble des couples d'amorces retenus présentent une efficacité théorique proche de 100% (Tableau 4).

L'ensemble des amorces retenues pour les expériences de PCR quantitative est rassemblé dans le tableau ci-dessous (Tableau 5)

**Tableau 5 Amorces utilisées pour la quantification de 24 transcrits par PCR quantitative en temps réel.**

La longueur des fragments amplifiés est indiquée en nucléotides dans la dernière colonne (nt).

Gène	No. Acc.	Oligonucleotide sens	Oligonucleotide antisens	nt
<i>AtRH1</i>	At4g15850	AGGTAAGGCGTTTCTCGAAGC	GAAGTAGGCGGGATCGGATAG	75
<i>AtRH2</i>	At3g19760	GAGGCGGTGGAGCTATGGA	CCCATATCATTGAAGCTCGTGA	84
<i>AtRH4</i>	At3g13920	GCTGTGTTTGCTCTGCCGTA	GATAGACGACGGGTCAAATCAG	93
<i>AtRH5</i>	At1g31970	CCGTCTCCGATTCAGTCACA	TGACCCAGTTTTTCGCAATCC	58
<i>AtRH6</i>	At2g45810	CAAGCCCATACTTCACTTATCG	ACACCGGTTAGTCCATCAATGC	100
<i>AtRH8</i>	At4g00660	GTTGAAGCCATGGTTAGACAGTTG	ATAAAGGAATCCCAAAGACGCAG	123
<i>AtRH12</i>	At3g61240	GTAACCCAACCCACTCAAACG	CACAGAAGAACGGAGAAGAAAGC	127
<i>AtRH14</i>	At3g01540	AGCTTCCACGAGACGATGATG	CCCACAACACAATCTTCTGTAC	89
<i>AtRH15</i>	At5g11170	GGGGGTTCACTATCTCTCG	TAAGTGTGCGATTTGGAGAGTG	127
<i>AtRH19</i>	At1g54270	CGTGGCTGATTTGCTGTGAAG	GTTTCGATAGAGTTGATGAAGGATACTAGAG	153
<i>AtRH20</i>	At1g55150	TTGGAGGATTCAGAGACCGTG	CTGAGCCCTAAAGCCATAACAAG	82
<i>AtRH21</i>	At2g33730	CGATAGGCTCTCTGGATTTGTTT	TGCTACTGCTTCTTCACTATCAAAC	129
<i>AtRH23</i>	At1g72730	GCTTTGATTTTTGTTTTTCAGGCTACTAC	TGCCGAAATTCCTTGATTGCC	106
<i>AtRH24</i>	At2g47330	GGAGACATGGGTAGAACTCAGAGTC	TTCTGGCTGGAGTTGTGTGATG	79
<i>AtRH24</i>	At2g47330	TCGGCATTAAATCGGCAAACATA	CGGCATCTTCGGAGTTGTCT	104
<i>AtRH24</i>	At2g47330	TAGAGAGGTACAAGGACGACGATG	CATCGGCGGCTAGTGTGACG	86
<i>AtRH30</i>	At5g63120	TCACGGTGGAAAGAGGACTTG	TCTAGTCGAAACCAAACGCTGA	172
<i>AtRH34</i>	At1g51380	GAAGCTGAAGTTATAGAATCGGTTTTG	GGTTGGTACAAAAGAAATGACTGAAG	150
<i>AtRH38</i>	At3g53110	AGTGAGGAAGAGAGACAGCATGTG	CTTAAAACAATCTCAAGAAAACAGACG	49
<i>AtRH40</i>	At3g06480	GTTGTTGAATCGGAGGTAGAAGC	GCTTGTGATATCTCCGTAGTGAGG	85
<i>AtRH42</i>	At1g20960	TCTGGTCTTTGTTTCAGTCGCAG	CCGTGAAGAGACAGACAAGGG	66
<i>AtRH56</i>	At5g11200	TCAAAGGGGTCATGGAGGTC	CTCATAGAAAAGTCTAAAATAGCGAAGC	74
<i>AtRH58</i>	At5g19210	TGTCGCTAATCTGATCACCTCG	CAACATCATTTCTCGCAGC	87
<i>AtACT2/8</i>	At3g18780 At1g49240	GGTAACATTGTGCTCAGTGGTGG	AACGACCTTAATCTTCATGCTGC	107
<i>AtEF1<math>\alpha</math>-4</i>	At5g60390	AGCGTGGACAAGAAGGACCC	TCAAAGTTTTAAGAACCATAAGAGAGTCG	115
<i>AtSK<math>\alpha</math></i>	At5g26751	CTTATCGGATTTCTCTATGTTTGGC	GAGCTCCTGTTTATTAACTTGACATAACC	130
<i>AtSUP1</i>	At3g23130	AGAAAGAGCTTGACATATGGAGAG	GGCCATGAAAACCCTAGAAGATAATC	144

#### IV. Validation de la PCR quantitative en temps réel dans le cadre de l'étude d'une famille de gènes

Du fait de leur origine commune, les gènes d'une famille présentent des séquences similaires, rendant difficile l'étude de leur expression par les approches de type *Northern Blot*. Celles-ci sont d'autant plus difficiles à réaliser que les gènes étudiés présentent des expressions faibles. La PCR quantitative en temps réel semble alors être un bon substitut aux techniques d'hybridation. En effet, la combinaison des oligonucléotides associée à l'amplification par PCR permettent d'une part une plus grande spécificité et d'autre part une plus grande sensibilité. Il est cependant nécessaire de déterminer le domaine d'application dans lequel la RT-PCR quantitative en temps réel est une approche valable, avant de l'appliquer à l'étude de la famille des *AtRH*.

Pour ce faire, l'étendue de la gamme d'ADN génomique permettant des mesures fiables et reproductibles, puis la limite de spécificité des amorces, la reproductibilité technique, la variabilité biologique et enfin la qualité des échantillons d'ADNc ont été évaluées.

##### A. Etendue de la gamme dynamique

En PCR quantitative, la qualité de la quantification d'un échantillon inconnu dépend directement de la qualité de la gamme-étalon. Il existe une partie de la courbe représentant le cycle-seuil (Ct) en fonction de la quantité initiale d'ADN ( $N_0$ ) où la variabilité expérimentale est la plus faible et où la fiabilité de la quantification est maximale. Les limites inférieure et supérieure de cette région ont été précisément déterminées par étapes successives.

Une première expérience nous a indiqué que la saturation est atteinte au-delà de 40 ng d'ADN génomique (soit 266 667 équivalents-génome) par puits. Cette saturation provient soit d'un excès d'ADN (inhibition par excès de substrat), soit de la présence d'inhibiteurs de PCR. Ensuite, en procédant par des dilutions successives, nous avons atteint la limite de 9 pg d'ADN génomique (soit 60 équivalents-génome) par puits sur une gamme de 12 points. L'amplification réalisée avec 17 équivalents-génome environ se confondait avec le bruit de fond. L'étendue maximale dans laquelle la gamme est linéaire et fiable est donc de 60 à 266 667 équivalents-génome, soit une amplitude de l'ordre de 4000 fois. Par la suite, nous avons limité l'étendue de la gamme aux valeurs du tableau ci-dessous (Tableau 6).

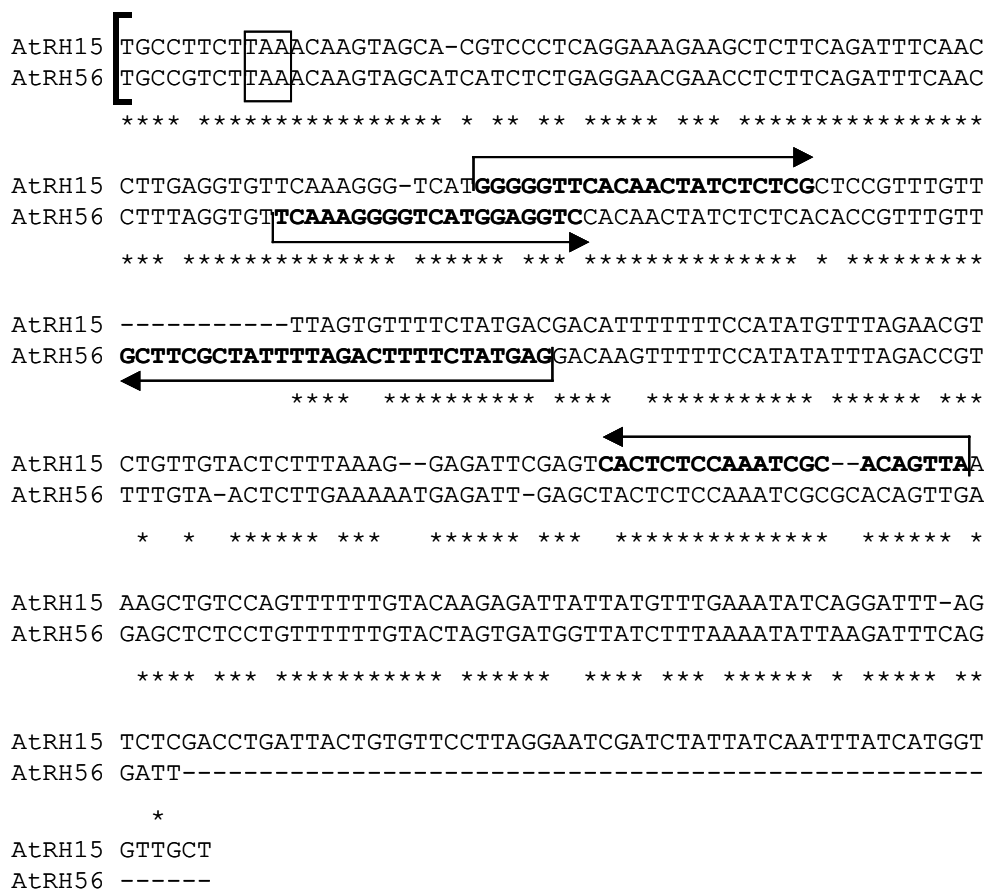
**Tableau 6** Correspondance entre la quantité d'ADN génomique et le nombre d'équivalents-génome chez *A. thaliana*.

Quantité d'ADN déposée (ng)	0,1	0,4	1,4	4,8	16,7
Equivalent-génome (eg)	740	2592	9070	31746	111111

**Tableau 7 Efficacité de la PCR des couples d'amorces dérivant du couple spécifique d'*AtRH15* par le dernier nucléotide.**

En gras sont signalés les nucléotides spécifiques d'*AtRH15*, en italique ceux spécifiques d'*AtRH56* et en minuscules les nucléotides n'existant ni dans *AtRH15*, ni dans *AtRH56*.

Amorces <i>AtRH15</i> antisens	Amorces <i>AtRH15</i> sens GGGGGTTCACAACTATCTCTCG			
	G	c	t	A
<b>TA</b> ACTGTGCGATTTGGAGAG <b>TG</b>	100,0%	0,3%	3,5%	23,3%
c	0,5%	0,0%	0,0%	0,8%
t	1,6%	3,4%	0,1%	0,3%
A	7,3%	0,4%	0,0%	20,1%



**Figure 5 Alignement du dernier exon des gènes *AtRH15* et *AtRH56*.**

La position des amorces est signalée en gras et leur orientation est indiquée par une flèche, située au-dessus de la séquence pour *AtRH15* et au-dessous de la séquence pour *AtRH56*. Les nucléotides identiques entre les deux gènes sont indiqués par les étoiles. L'accolade symbolise le site d'épissage et le début de l'exon. Le codon STOP est encadré, rappelle la fin de la région codante et le début de la région 3' UTR des deux gènes.

## B. Spécificité des amorces

L'étude de l'expression d'une famille de gènes par PCR quantitative pose le problème de discrimination des différents membres de la famille, dont les séquences sont proches, par des couples d'amorces spécifiques de chaque gène. La famille des *AtRH* comporte des gènes récemment dupliqués et très proches en séquence. Nous avons cherché à déterminer les limites de la spécificité de la PCR quantitative.

Nous nous sommes tout d'abord intéressés au cas des gènes *AtRH4* et *AtRH19*. Leurs séquences présentent 88% d'identité et divergent surtout au niveau des extrémités non codantes. Le couple d'oligonucléotides défini pour chaque gène est composé d'une amorce spécifique du gène, présentant plus de 50% de différence entre le gène ciblé et son paralogue, et d'une autre amorce, ne divergeant que par quelques nucléotides et pouvant s'hybrider potentiellement sur les deux gènes. La spécificité de ces couples d'amorces a été établie par l'amplification d'un seul produit correspondant au gène ciblé. Ceci a été vérifié par l'obtention d'un pic unique après réalisation de la courbe de fusion des produits PCR, par l'observation d'une bande unique sur gel coloré au BEt et par séquençage. La présence d'une seule amorce spécifique du gène considéré suffit à assurer l'amplification du gène désiré uniquement. Par ailleurs, la suppression du dernier nucléotide des amorces influence peu l'efficacité de la PCR, même si une plus grande variabilité de l'efficacité de l'amplification a été observée.

Les transcrits des gènes *AtRH15* et *AtRH56* présentent plus de 97% d'identité au niveau de la partie codante et 74% d'identité dans la partie la plus divergente (3' UTR). Les seules amorces qu'il a été possible de définir pour le gène *AtRH15* divergent de la séquence d'*AtRH56* par uniquement 2 à 3 nucléotides dont un seul à l'extrémité 3' (Figure 5). Pour tester la spécificité de ces amorces, nous avons mesuré l'influence du mésappariement du dernier nucléotide des amorces sur l'efficacité de la PCR. Le dernier nucléotide de chaque amorce correspondant au gène *AtRH15* est un G, substitué en T dans la séquence du gène *AtRH56*. A partir des amorces du gène *AtRH15*, six autres amorces ont été générées en échangeant la dernière base (G) contre un A, un C ou un T. L'efficacité des 16 expériences de PCR quantitative testant l'ensemble des combinaisons possibles des huit amorces différant uniquement par le dernier nucléotide sont présentées ci-contre (Tableau 7).

Lorsqu'un seul des oligonucléotides présente un mésappariement avec la matrice, l'efficacité de la PCR est considérablement réduite : un mésappariement purine-pyrimidine (G/A ou T) entraîne une inhibition de 96,5% à 98,4% et un mésappariement G/C une inhibition supérieure à 99,5%. Lorsque les deux amorces présentent un mésappariement avec la matrice, l'inhibition est totale dans 3 cas sur 4 et de 99,5% pour le 4ème cas. Lorsque la dernière base de l'amorce correspond à la séquence nucléotidique de l'autre paralogue l'inhibition de la PCR est variable : de 77% à 93% environ. Nous avons également observé ce phénomène avec le gène *AtRH56* dont un des oligonucléotides est spécifique et l'autre ne diverge de la séquence d'*AtRH15* que par trois nucléotides, dont un seul à l'extrémité 3'.

Dans les puits où la quantité de produit PCR était significative, nous avons vérifié la courbe de fusion : dans tous les cas nous n'avons observé qu'un seul pic correspondant au produit spécifique attendu. Ainsi, nous avons pu montrer que, dans nos conditions expérimentales, le dernier nucléotide de chaque amorce est suffisant pour assurer la spécificité de la PCR.

### C. Stratégie expérimentale

Etudier l'expression d'une famille de gènes par RT-PCR quantitative en temps réel implique des choix expérimentaux spécifiques qui prennent en considération le nombre important de gènes mesurés. Le nombre de manipulations du protocole standard a donc été réduit au maximum pour limiter d'autant les sources d'erreurs, le temps nécessaire pour chaque mesure et le coût total des expériences.

Tout d'abord, j'ai séparé l'étape de transcription inverse (RT) de l'étape de PCR quantitative car cela permet de mesurer les teneurs de plusieurs transcrits à partir des ADNc d'une seule réaction de RT. Pour ce faire, les ADNc simple brin ont été synthétisés à l'aide d'oligo(dT) pour convertir l'ensemble des ARNm polyadénylés en ADNc. La synthèse du second brin, nécessitant un oligonucléotide spécifique, n'est pas réalisée. Ce système dissocie l'étape de RT de l'étape de PCR : il permet de réduire le nombre de RT et donc la variation introduite par la RT dans la mesure finale.

Pour la PCR quantitative en temps réel, j'ai trouvé plus intéressant d'utiliser le SYBR-Green® comme molécule fluorogène plutôt que des sondes oligonucléotidiques. Ce choix est motivé par le nombre de gènes étudiés et par leur expression a priori faible. Tout d'abord, cette solution est économique. En effet, l'utilisation de sondes implique la conception d'un troisième oligonucléotide par gène étudié. Or, ces sondes sont marquées par des fluorophores

et coûtent plus chers que des oligonucléotides classiques. L'aspect financier mis à part, l'utilisation d'un troisième oligonucléotide augmente le nombre de manipulations à chaque PCR et le nombre de tubes à stocker. Le SYBR-Green® permet un marquage du fragment d'ADN amplifié, indépendamment de sa séquence. Ainsi, cette molécule peut être vue comme un marqueur fluorescent « universel », contrairement aux sondes qui sont spécifiques du produit amplifié. Mais l'avantage principal du SYBR-Green® est la grande sensibilité qu'il confère à la mesure. Comme chaque fragment d'ADN amplifié est marqué par plusieurs molécules de SYBR-Green®, le signal qu'il génère est plus intense que celui des sondes.

Le SYBR-Green® a donc été retenu comme agent fluorescent unique, utilisable pour tous les transcrits. Pour assurer la simplicité du protocole et garantir une reproductibilité maximale des mesures, nous avons choisi de travailler avec un mélange commercial 2X « prêt-à-l'emploi », contenant l'enzyme, les dNTP et le SYBR-Green®, auquel il suffit d'ajouter les deux amorces et l'eau.

#### D. Reproductibilité technique et biologique

La méthode permettant de mesurer l'expression transcriptionnelle d'un gène par RT-PCR nécessite trois étapes principales : tout d'abord, les ARN totaux sont extraits de différents organes végétaux, puis ils sont convertis en ADNc par transcription inverse et enfin, l'abondance d'un transcrit dans un échantillon d'ADNc est estimée par PCR quantitative en temps réel. Ces différentes étapes introduisent chacune une variabilité sur la mesure finale (Figure 6A). Pour estimer l'impact de ces trois étapes sur la variabilité des mesures, nous avons comparé l'expression transcriptionnelle de 20 gènes *AtRH* et de 3 autres gènes, *EF1 alpha 4*, *ACT2/8* et *SUPERMAN (SUP1)*, par PCR quantitative en temps réel dans différents organes d'*A. thaliana* : racines, feuilles de la rosette jeunes et âgées, hampes florales, feuilles caulines jeunes et âgées, bourgeons floraux, fleurs et jeunes siliques.

Les organes végétaux cités ci-dessus ont été prélevés séparément sur plusieurs lots de plantes indépendants pour constituer 27 échantillons biologiques. Chaque organe est représenté par 3 échantillons, sauf les jeunes feuilles caulines et les tiges des hampes florales, représentées deux fois et, les racines et les siliques, représentées 4 fois. Les ARN totaux ont été extraits de ces échantillons indépendants. Plusieurs extractions ayant pu être réalisées à partir du même

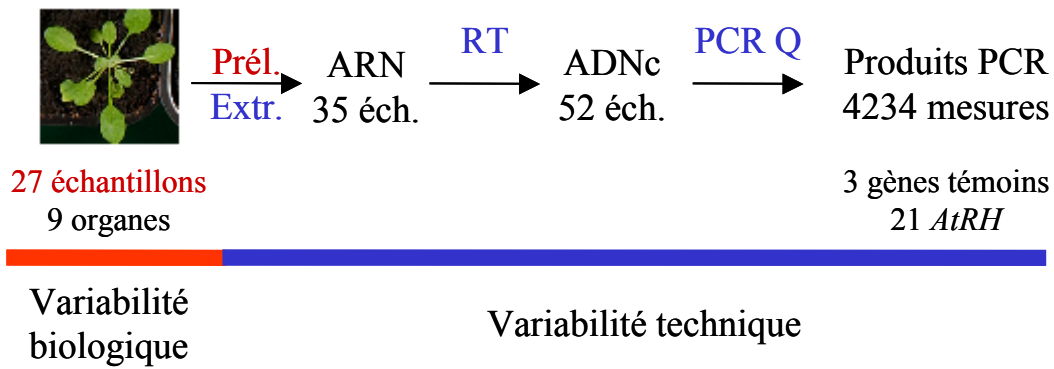


Figure 6A Schéma de synthèse des mesures de l'expression de 24 gènes par RT-PCR quantitative en temps réel.

A chaque étape, le nombre d'échantillons obtenus et la nature de la variabilité introduite ont été précisés. Prél. : Prélèvement des échantillons végétaux ; Extr. : Extraction des ARN totaux ; PCR Q : PCR quantitative en temps réel ; Ech. : Echantillons.

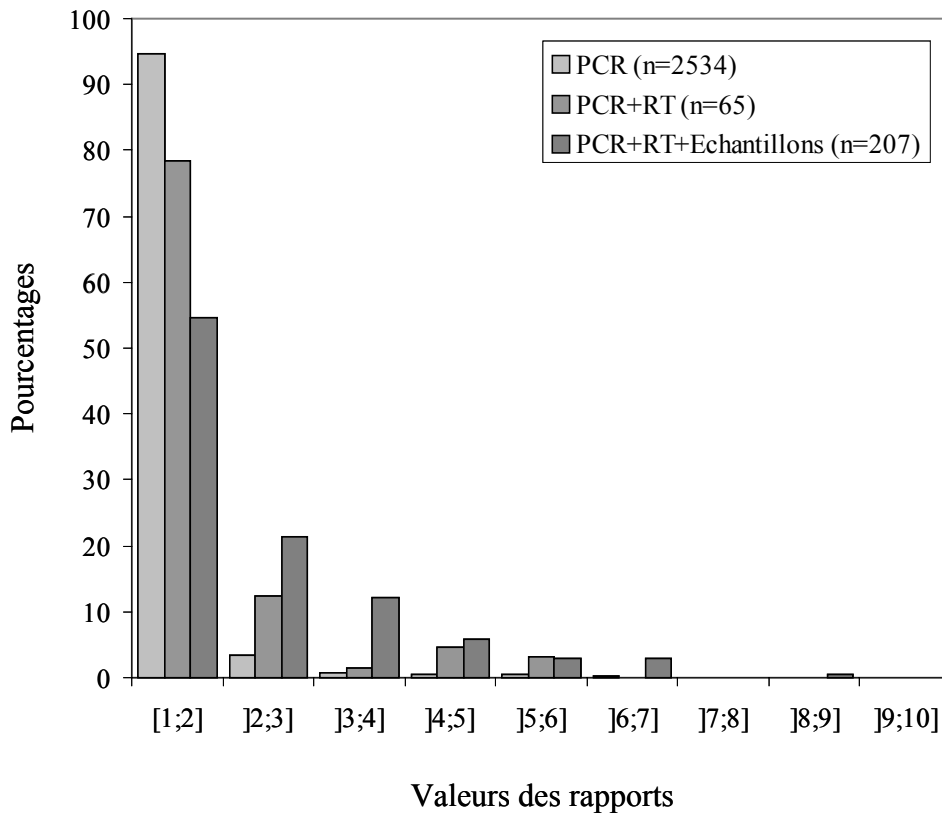


Figure 6B Variabilité des mesures de l'abondance des transcrits de 20 gènes *AtRH*, du gène *EF1 alpha* et des gènes *ACT2/ACT8* mesurée par PCR quantitative en temps réel dans 9 organes différents d'*A. thaliana*.

Les mesures de PCR quantitative en temps réel ont été comparées entre elles pour évaluer l'impact de chaque étape du protocole sur la variabilité totale des résultats. Les rapports obtenus indiquent la variabilité introduite par la PCR (gris clair), celle introduite par la RT et la PCR (gris moyen) et la variabilité totale (gris foncé) des résultats. Ils ont été répartis en classes selon leur valeur. Le nombre total de comparaisons (n) réalisées pour chaque étape est indiqué entre parenthèses sur le graphique.



échantillon, 35 stocks indépendants d'ARN totaux ont été obtenus. Ces ARN totaux ont été convertis en ADNc par transcription inverse. Lorsque la quantité d'ARN d'un stock le permettait, l'étape de RT a été répétée jusqu'à 4 fois. Les 52 échantillons d'ADNc simple brin ainsi obtenus ont été utilisés directement comme matrice pour mesurer l'abondance des transcrits spécifiques des 24 gènes dans les différents organes, après une dilution adéquate (Figure 6A).

La variabilité introduite par l'étape de PCR quantitative, par l'étape de transcription inverse et par la préparation des échantillons biologiques a été évaluée en comparant deux à deux les mesures. Les rapports ainsi obtenus ont été répartis en différentes classes (Figure 6B).

La reproductibilité de l'étape de PCR quantitative a été évaluée en comparant les mesures de l'expression transcriptionnelle d'un gène réalisées sur le même échantillon d'ADNc. Ainsi, 2534 comparaisons ont pu être réalisées avec les données de 23 gènes différents. 2396 rapports (95%) sont compris entre 1 et 2 et 2485 (98%) entre 1 et 3. La variabilité introduite uniquement par l'étape d'amplification par PCR est donc très faible. En ce qui concerne les 49 rapports restants (1%), dont les valeurs sont supérieures à 3, le maximum atteint est de 7,8. Les données correspondantes ont été écartées de l'analyse finale.

Concernant la variabilité introduite par l'étape de transcription inverse, les mesures de l'abondance des transcrits sur les ADNc provenant de RT différentes réalisées sur le même stock d'ARN totaux ont été comparées. Les mesures de l'expression transcriptionnelle de 9 gènes dans différents organes ont permis d'établir 65 rapports différents. Leur répartition se révèle bimodale. 91% des rapports sont inférieurs ou égaux à 3 mais on observe une seconde population (6 rapports soit 9%) formant un pic centré autour de la valeur 4,5. La valeur maximale des rapports atteint 5,8. Ce pic correspond aux quelques valeurs aberrantes déjà présentes lors de l'étude de la marge d'erreur introduite par la PCR. Il n'est donc pas dû à un effet de la reproduction de l'étape de RT. Si l'on retire ces valeurs de l'analyse, 85% des rapports sont compris entre 1 et 2.

La variabilité associée à l'extraction n'a pas été estimée séparément de la variabilité biologique. Elles ont été estimées à partir l'expression transcriptionnelle moyenne de chaque gène dans les différents organes, les mesures obtenues à partir d'extractions réalisées en parallèle sur le même échantillon végétal ayant été rassemblées. L'ensemble des gènes n'a pas

été étudié dans l'ensemble des échantillons biologiques. Ainsi, 207 rapports ont pu être calculés à partir des données de 16 gènes et de 27 échantillons biologiques. 157 rapports (76%) sont compris entre 1 et 3. Les 50 rapports supérieurs à 3 se répartissent de façon décroissante jusqu'à la valeur 9,0. Ces rapports supérieurs à 3 proviennent de mesures effectuées essentiellement dans deux organes : les hampes florales et les feuilles caulines. Cela pourrait provenir d'une plus grande variabilité du développement de ces organes, rendant difficile l'estimation de l'état de différenciation des tissus et donc l'homogénéité des prélèvements.

Nous observons donc que lorsque l'expression transcriptionnelle d'un gène est estimée à partir des transcrits provenant d'une seule extraction, il y a 95% de chance que les mesures ne varient que d'un facteur 2. Par contre, lorsqu'on compare deux échantillons biologiques, on observe que 75% des mesures varient dans une gamme de 1 à 3. La marge d'erreur introduite par l'obtention des transcrits est supérieure à celle introduite par les réactions enzymatiques des deux étapes suivantes mais elle reste largement inférieure à la variabilité biologique. La variabilité technique étant inférieure à la variabilité biologique, l'ensemble de ce protocole permet donc de mesurer de façon fiable l'expression transcriptionnelle des gènes chez *A. thaliana*. De plus, nous pouvons affirmer, avec un seuil de confiance de 95%, que la variation d'un facteur 2 observée en routine entre deux mesures par PCR quantitative est non significative. En conséquence de cette analyse, nous avons décidé de mener notre étude comparative de l'expression transcriptionnelle de 20 gènes *AtRH* en utilisant un seul lot de plantes afin de nous affranchir de la variabilité introduite par la préparation de l'échantillon. De plus, pour cette étude, le niveau de transcrits d'*AtRH4* a été utilisé pour détecter les éventuels problèmes au cours de l'étape de PCR et pour permettre la standardisation des données et faciliter leur représentation.

#### E. Qualité des échantillons

Les racines, les feuilles de la rosette jeunes et âgées, les hampes florales, les feuilles caulines jeunes et âgées, les bourgeons floraux, les fleurs et les jeunes siliques ont été prélevés sur 48 plants d'*A. thaliana*. Les ARN totaux ont été extraits de ces 9 organes et convertis en ADNc. Pour vérifier la qualité de ces 9 échantillons d'ADNc, plusieurs contrôles ont été réalisés par PCR quantitative en temps réel.



## **V. Conclusion**

Au cours de ma thèse, j'ai pu démontrer tout d'abord que la gamme dynamique des mesures des teneurs en transcrits par RT-PCR quantitative est très large : son amplitude maximale est de l'ordre d'un facteur 4000. En routine, la gamme utilisée permet de quantifier une différence de quantité d'un transcrit entre deux échantillons d'un facteur 150. D'autre part, la mesure de l'abondance des transcrits par RT-PCR quantitative est fiable et reproductible. La variation introduite dans les résultats par cette technique est largement inférieure à celle introduite par la variabilité biologique. Typiquement, la variation des mesures observée en routine est de l'ordre d'un facteur 2. De plus, la PCR quantitative permet de mesurer le niveau des transcrits et d'établir le profil d'expression transcriptionnel de gènes dont l'expression varie dans une large gamme d'intensité et de répartition selon les organes. La RT-PCR quantitative en temps réel apparaît comme l'outil le plus approprié pour l'analyse de l'expression transcriptionnelle d'une famille de gènes.

## **Expression des *AtRH* et analyse des promoteurs**

Les résultats exposés dans cette partie font l'objet d'une publication acceptée dans la revue « *Plant Biotechnology Journal* ».

**Mingam A, Toffano-Nioche C, Brunaud V, Boudet N, Kreis M and Lecharny A.**

DEAD-box RNA helicases in *Arabidopsis thaliana*: establishing a link between quantitative expression, gene structure and evolution of a family of genes.

## I. Introduction

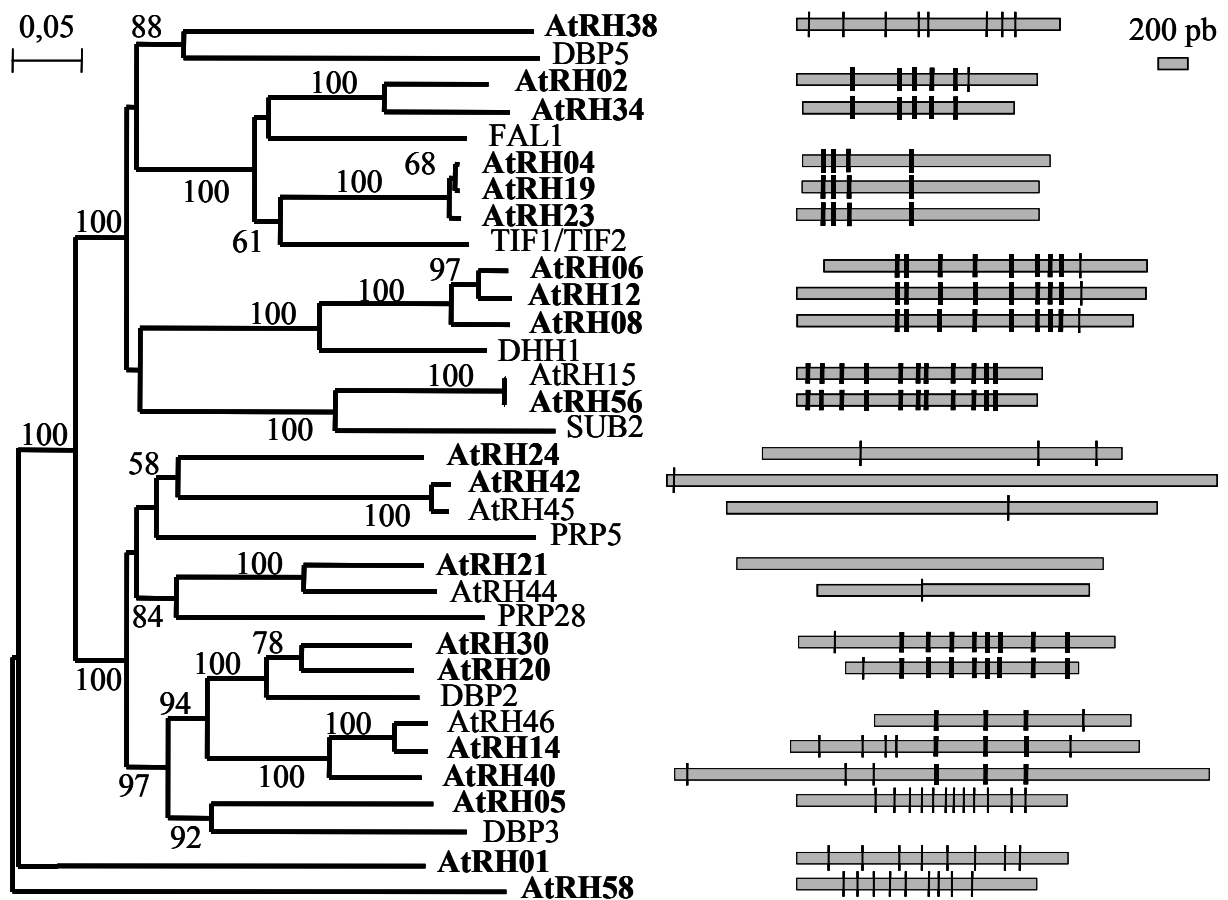
La famille des ARN hélicases à boîte DEAD (RH) appartient à la superfamille des ARN hélicases non virales (SFII) (Schmid & Linder, 1992; Lüking *et al.*, 1998; De la Cruz *et al.*, 1999). Participant à la transcription, à la traduction, à l'épissage des ARN et à l'assemblage des ribosomes, les membres de la famille des ARN hélicases à boîte DEAD sont considérés comme des gènes « de maintenance ». Cependant, certaines RH ont aussi été impliquées dans des processus plus spécifiques au cours du développement et de la différenciation cellulaire. Les familles des *RH* d'*A. thaliana*, *C. elegans* et *D. melanogaster* contiennent respectivement 58, 32 et 30 gènes (Boudet *et al.*, 2001), tandis que le génome nucléaire de *S. cerevisiae* en contient 26. Une étude exhaustive de toutes les protéines impliquées dans le métabolisme des ARN, comparant les génomes complets bactériens, archaebactériens et eucaryotes, confirme l'importante prolifération de la superfamille *SFII* chez les eucaryotes (Anantharaman *et al.*, 2002). De plus, comme pour la famille des *RH*, l'expansion de la superfamille *SFII* la plus prononcée est observée dans le génome d'*A. thaliana*. En effet, ce génome contient 103 gènes *SFII* alors que les autres génomes eucaryotes entièrement séquencés, dont celui de l'Homme, possèdent environ 50 *SFII* gènes chacun.

Cette augmentation importante du nombre de gènes *SFII* et *RH* chez les plantes peut être corrélée au recrutement de certains gènes dupliqués pour assurer de nouvelles fonctions. Rechercher des changements de spécificité tissulaire ou temporelle de l'expression est un moyen de démontrer la spécialisation de gènes dupliqués. Et, mettre en regard les données d'expression des gènes avec les arbres phylogéniques donne des informations sur les différents mécanismes de spécialisations envisageables comme la sous-fonctionnalisation ou la néo-fonctionnalisation par exemple. Aucune étude exhaustive de l'expression des RH d'un génome eucaryote pluricellulaire n'a encore été réalisée, mais des *RH* présentant une expression spécifique pendant le cycle cellulaire, le développement ou sous l'effet d'un stress ont été identifiées chez les animaux (De Valoir *et al.*, 1991; Russell & Bennett, 1993; Fujiwara *et al.*, 1994; Eberl *et al.*, 1997; Fujimura & Takamura, 2000; MacArthur *et al.*, 2000; Navarro *et al.*, 2001) et les végétaux (Brander *et al.*, 1995; Zegzouti *et al.*, 1999). Au premier abord, il est surprenant de trouver des spécificités d'expression pour des gènes assurant des fonctions de base du métabolisme cellulaire. Ce paradoxe est d'ailleurs soulevé dans plusieurs discussions sur la normalisation des données d'expression issues de méthodes d'analyse à haut débit (Pfaffl, 2001; Vandesompele *et al.*, 2002) et la PCR quantitative en temps réel (Bustin, 2002). Ces résultats suggèrent que, pendant l'évolution des eucaryotes supérieurs, la

fonction des gènes « de maintenance cellulaire » (par exemple les *RH*) s'est répartie entre les gènes dupliqués par une spécialisation développementale de l'expression et/ou le recrutement de quelques paralogues pour des fonctions différentes des fonctions de base.

Chez les plantes supérieures, les mécanismes de régulation transcriptionnelle sont très étudiés chez les gènes présentant une régulation spatiale et temporelle de leur profil d'expression. Récemment, des études par PCR quantitative en temps réel ont été utilisées pour rechercher des spécificités d'expression prononcées dans quelques familles de gènes (Yokoyama & Nishitani, 2001; Charrier *et al.*, 2002; Orsel *et al.*, 2002b). De plus, les analyses de gènes végétaux fortement exprimés ont permis d'identifier des éléments de l'architecture des gènes potentiellement impliqués dans la détermination de leur fort niveau d'expression (Sawant *et al.*, 2001; Tremousaygue *et al.*, 2003). Nous avons choisi une approche différente en nous intéressant aux différences de niveau transcriptionnel entre des gènes paralogues au lieu de gènes fortement exprimés sans liens fonctionnels. Ce travail s'articule autour de trois questions : 1) Est-ce que la PCR quantitative en temps réel est adaptée pour comparer qualitativement et quantitativement les profils transcriptionnels ? Dans une revue récente, Bustin (2002) commente combien il est difficile de répondre à cette question avec les données disponibles. J'ai évalué la méthode de PCR quantitative en temps réel par l'analyse des profils transcriptionnels et des niveaux d'expression de 20 *AtRH* dans neuf organes d'*A. thaliana*. 2) Est-ce que certains paralogues *AtRH* sont recrutés pour effectuer des fonctions spécifiques ? Pour cela, j'ai comparé le profil d'expression de ces 20 *AtRH* dans différents organes. 3) Quels sont les éléments qui participent au contrôle du niveau d'expression transcriptionnelle des *AtRH* ? Afin de tenter d'expliquer la large gamme de niveaux de transcription des *AtRH*, les données d'expression ont été combinées à l'organisation des structures *cis*-régulatrices des *AtRH*.

Les résultats obtenus pour répondre à ces trois questions m'ont permis de comparer l'évolution des séquences protéiques, de la structure des gènes et de l'activité transcriptionnelle de plusieurs groupes fonctionnels d'*AtRH*.



**Figure 7** Arbre de distance d'ARN hélicases à boîte DEAD d'*A. thaliana* et de *S. cerevisiae* et structures intron/exon des gènes *AtRH* (d'après Boudet, 2002).

L'arbre de distance est basé sur l'alignement des motifs protéiques conservés des RH entre *A. thaliana* et *S. cerevisiae*. Il comprend l'ensemble des *AtRH* étudiées au cours de ce travail et leurs paralogues proches associées aux ScRH correspondantes. La valeur des *bootstraps* a été calculée à partir de 100 répétitions. Les noms en gras indiquent les gènes sélectionnés pour l'étude de leur expression par PCR quantitative en temps réel.

Les exons des *AtRH* sont représentés par les rectangles gris et les introns par des traits verticaux. En gras, les introns à une position conservée ; en traits fins, les introns non conservés (Boudet, 2002).



## II. Expression transcriptionnelle de 20 *AtRH*

### A. Choix des gènes

L'ensemble des gènes de la famille des RH a été identifié chez *A. thaliana*, *C. elegans* et *D. melanogaster* par analyse bioinformatique au laboratoire (Boudet *et al.*, 2001). Leurs séquences protéiques ont été utilisées avec celles des gènes de *S. cerevisiae* pour construire un arbre de distance. Cet arbre a permis de définir des groupes associant fortement des gènes des 4 organismes. Comme cela a déjà été observé pour d'autres familles de gènes (Lecharny *et al.*, 2003), les *AtRH* présentent une grande divergence de leur structure intron-exon : le nombre d'introns varie de 0 à 18 selon le paralogue et il existe 28 structures de gène différentes (Boudet *et al.*, 2001). Cette divergence structurale des *AtRH* vient renforcer les branches terminales de l'arbre. En effet, la structure des *AtRH* appartenant à un même groupe d'orthologues présumés est le plus souvent identique (Boudet *et al.*, 2001). A partir des branches terminales de l'arbre, 18 *AtRH* appartenant à 9 groupes de gènes, représentatifs des différentes fonctions des RH identifiées chez *S. cerevisiae* (De la Cruz *et al.*, 1999; Tanner & Linder, 2001) ont été retenus afin d'établir leur profil d'expression transcriptionnelle dans différents organes d'*A. thaliana* (Figure 7). L'expression transcriptionnelle a été mesurée pour les gènes suivants : les gènes *AtRH4*, *19* et *23* associés aux gènes de la levure *ScRH-TIF1* et *-TIF2*, impliqués dans l'initiation de la traduction; les gènes *AtRH6*, *8* et *12* associés au gène *ScRH-DHH1* impliqué dans la transcription (Coller *et al.*, 2001 ; Fischer & Weis, 2002) ; le gène *AtRH21* associés au gène *ScRH-PRP28*, impliqué dans l'épissage des ARNm ; les gènes *AtRH24* et *42* associés au gène *ScRH-PRP5*, impliqué dans l'épissage des ARNm; le gène *AtRH56* associé au gène *ScRH-SUB2*, impliqué à la fois dans l'épissage des ARNm et leur export vers le cytoplasme (Jensen *et al.*, 2001; Linder & Stutz, 2001; Strasser & Hurt, 2001) ; le gène *AtRH5* associé au gène *ScRH-DBP3*, impliqué dans la biogenèse des ribosomes 60S (Weaver *et al.*, 1997) ; les gènes *AtRH2* et *34* associés au gène *ScRH-FAL1*, impliqué dans les premières étapes de la maturation des ARNr 35S dans le nucléole (Kressler *et al.*, 1997) ; les gènes *AtRH14*, *20*, *30* et *40* associés au gène *ScRH-DBP2*, impliqué dans la biogenèse des ribosomes et la dégradation des ARN non-sens (Bond *et al.*, 2001) et le gène *AtRH38* associé au gène *ScRH-DBP5* impliqué dans l'export des ARNm vers le cytoplasme (Zhao *et al.*, 2002). Deux gènes isolés dans l'arbre, *AtRH1* et *AtRH58*, ont été également étudiés. Aucun orthologue à ces deux gènes n'est identifiable chez *S. cerevisiae*, *C. elegans* ni *D. melanogaster*. Par ailleurs, l'expression des gènes *AtRH44*, *45* et *46*, associés respectivement aux gènes *ScRH-PRP28*, *-PRP5* et *-DBP2*, n'a pas été mesurée. En effet, l'absence d'EST disponibles pour ces 3 gènes au moment des choix expérimentaux laissait

Ctrl -	Ctrl -	ADNg	ADNg	ADNg	ADNg	ADNg	ADNg	ADNg	ADNg	ADNg	ADNg
FJ-20	FJ-20	FJ-4	FJ-4	FJ-1	FJ-1	VF-20	VF-20	VF-4	VF-4	VF-1	VF-1
JC-20	JC-20	JC-4	JC-4	JC-1	JC-1	FC-20	FC-20	FC-4	FC-4	FC-1	FC-1
BF-20	BF-20	BF-4	BF-4	BF-1	BF-1	Fl-20	Fl-20	Fl-4	Fl-4	Fl-1	Fl-1
R-20	R-20	R-4	R-4	R-1	R-1	T-20	T-20	T-4	T-4	T-1	T-1
Si-20	Si-20	Si-4	Si-4	Si-1	Si-1	<b>Ctrl -</b>	<b>Ctrl -</b>	<b>ADNg</b>	<b>ADNg</b>	<b>ADNg</b>	<b>ADNg</b>
<b>ADNg</b>	<b>ADNg</b>	<b>ADNg</b>	<b>ADNg</b>	<b>ADNg</b>	<b>ADNg</b>	<b>FJ-1</b>	<b>FJ-1</b>	<b>VF-1</b>	<b>VF-1</b>	<b>JC-1</b>	<b>JC-1</b>
<b>FC-1</b>	<b>FC-1</b>	<b>BF-1</b>	<b>BF-1</b>	<b>Fl-1</b>	<b>Fl-1</b>	<b>R-1</b>	<b>R-1</b>	<b>Si-1</b>	<b>Si-1</b>	<b>T-1</b>	<b>T-1</b>

**Figure 8** Plan de répartition des échantillons d'une expérience de PCR quantitative en temps réel dans une plaque de 96 puits.

Chaque puits est symbolisé par une case du tableau. La nature de la matrice est indiquée par les abréviations suivantes : Ctrl -, contrôle négatif (eau) ; ADNg, gamme d'ADN génomique d'*A. thaliana* ; FJ, jeunes feuilles de la rosette ; VF, vieilles feuilles de la rosette ; JC, jeunes feuilles caulinaires ; FC, feuilles caulinaires matures ; BF, boutons floraux ; Fl, fleurs ; R, racines ; Si, jeunes siliques ; et T, tiges des hampes florales. La quantité d'ADNc simple brin déposée, en nanogramme d'équivalents ARN totaux, est indiquée par les chiffres suivant les abréviations. L'exemple présenté correspond à un gène dont l'expression moyenne est comprise entre  $10^3$  et  $5.10^3$  eg.ng ARN totaux<sup>-1</sup>. Les 66 premiers puits sont attribués au gène étudié et les 30 derniers puits (en gras) au gène *AtRH4*.

supposer un niveau d'expression faible. De fait, les régions non traduites n'étant pas expérimentalement connues, il n'a pas été possible de positionner des amorces spécifiques.

## B. Niveaux d'expression transcriptionnelle

### 1. PCR quantitative en temps réel

D'après les résultats du chapitre précédent, nous nous sommes placés dans les meilleures conditions expérimentales de PCR quantitative en temps réel pour comparer l'expression transcriptionnelle de 20 gènes *AtRH* dans différents organes d'*A. thaliana*. Pour mesurer l'abondance des transcrits dans des échantillons homogènes et statistiquement valides, les organes ont été prélevés à partir d'un même lot de 96 plantes. Elles ont été tout d'abord cultivées en jours courts pour favoriser leur développement végétatif jusqu'au stade « rosette 8 à 12 feuilles ». A ce stade, les racines (R), les jeunes feuilles de la rosette (FJ) et les feuilles âgées (VF) de 48 plantes ont été prélevées. Les 48 plantes restantes ont été transférées en jours longs pour induire la transition florale. Les jeunes feuilles autour des boutons (JC), les feuilles caulines (FC), les boutons floraux (BF), les fleurs (FI), les jeunes siliques (Si) et les tiges des hampes florales (T) ont été prélevés au bout d'un mois de croissance dans ces conditions. L'extraction des ARN et leur conversion en ADNc simple brin par transcription inverse ont été réalisées à la suite. L'abondance des transcrits a ensuite été mesurée par PCR quantitative dans les échantillons ainsi obtenus.

Pour garantir la validité des mesures par PCR quantitative en temps réel, nous avons déposé en duplicata pour chaque gène, trois dilutions sériées<sup>11</sup> de chaque échantillon. Ainsi, il est possible de vérifier la linéarité des mesures, c'est-à-dire de s'assurer que les mesures appartiennent à une zone comprise entre le bruit de fond et de la saturation. De plus, pour détecter un éventuel problème survenu au cours de la réaction de PCR, nous mesurons parallèlement sur la même plaque la teneur des transcrits du gène *AtRHA*, dans tous les échantillons (Figure 8).

L'abondance des transcrits dans neuf organes d'*A. thaliana* des 20 gènes *AtRH* cités ci-dessus a été mesurée par PCR quantitative (Tableau 9). Elle est indiquée par la moyenne des valeurs appartenant à la zone linéaire de la mesure c'est-à-dire 4 à 6 points dans la plupart des cas, plus ou moins l'écart-type. L'abondance des transcrits du gène *AtRHA*, utilisée comme

---

<sup>11</sup> Un facteur de 4 à 5 est recherché entre chaque dilution ; les quantités d'échantillon déposées suivent le plus souvent la gamme suivante : 80-20-5-1-0,2-0,05 ng d'équivalent ARN totaux.

**Tableau 9** Abondance des transcrits de 20 gènes *AtRH* et du gène *AtEF1α-4* dans neuf organes d'*A. thaliana*, mesurée par RT-PCR quantitative en temps réel.

R, racines; JF, jeunes feuilles de la rosette; VF, vieilles feuilles de la rosette; JC, jeunes feuilles caulines; FC, feuilles caulines matures; BF, boutons floraux; FI, fleurs; Si, jeunes siliques; et T, tiges des hampes florales. L'abondance des transcrits est exprimée en équivalent-génomomes par nanogramme d'ARN totaux.

	R	FJ	VF	FC	JC	BF	FI	Si	T	Total
<i>AtRH1</i>	172 ± 4	97 ± 1	52 ± 2	39 ± 2	48 ± 1	139 ± 10	102 ± 3	47 ± 5	46 ± 3	742
<i>AtRH2</i>	3763 ± 381	2157 ± 313	1231 ± 169	957 ± 88	1236 ± 146	3663 ± 907	2671 ± 161	1585 ± 110	970 ± 122	18233
<i>AtRH4</i>	18679 ± 2785	12171 ± 1856	7609 ± 1791	5041 ± 798	6587 ± 1114	10313 ± 2105	7048 ± 1920	5834 ± 1032	3261 ± 798	76543
<i>AtRH5</i>	1345 ± 193	978 ± 100	722 ± 80	390 ± 13	577 ± 71	1580 ± 242	1123 ± 178	449 ± 36	362 ± 50	7526
<i>AtRH6</i>	1464 ± 313	1132 ± 112	852 ± 55	1081 ± 227	1076 ± 79	1951 ± 602	1830 ± 310	890 ± 133	689 ± 179	10965
<i>AtRH8</i>	1951 ± 201	964 ± 109	586 ± 60	392 ± 7	538 ± 66	1475 ± 382	1151 ± 181	743 ± 29	345 ± 69	8145
<i>AtRH12</i>	1339 ± 384	705 ± 162	435 ± 89	331 ± 46	452 ± 16	1279 ± 300	861 ± 205	574 ± 117	251 ± 46	6227
<i>AtRH14</i>	6822 ± 456	3883 ± 477	3399 ± 122	3024 ± 1104	2984 ± 657	4426 ± 1042	5933 ± 922	1692 ± 307	2900 ± 249	35063
<i>AtRH19</i>	7549 ± 1722	7704 ± 1351	3476 ± 460	3814 ± 1032	3959 ± 53	8798 ± 946	8755 ± 334	3366 ± 631	2727 ± 36	50148
<i>AtRH20</i>	1247 ± 109	602 ± 24	369 ± 46	395 ± 30	394 ± 44	994 ± 186	897 ± 85	356 ± 30	456 ± 34	5710
<i>AtRH21</i>	2547 ± 389	1715 ± 122	946 ± 83	1067 ± 192	1105 ± 80	2534 ± 859	2139 ± 389	1090 ± 128	703 ± 103	13846
<i>AtRH23</i>	1346 ± 207	767 ± 71	230 ± 31	77 ± 4	210 ± 7	825 ± 213	499 ± 71	393 ± 18	305 ± 46	4652
<i>AtRH24</i>	881 ± 131	462 ± 52	320 ± 38	400 ± 57	368 ± 37	758 ± 193	679 ± 98	357 ± 36	275 ± 42	4500
<i>AtRH30</i>	1110 ± 77	507 ± 52	250 ± 89	560 ± 34	595 ± 95	1099 ± 194	1017 ± 78	430 ± 71	454 ± 30	6022
<i>AtRH34</i>	389 ± 18	76 ± 5	26 ± 2	23 ± 1	48 ± 4	146 ± 20	125 ± 19	40 ± 5	50 ± 6	923
<i>AtRH38</i>	1824 ± 165	1252 ± 195	992 ± 16	639 ± 195	813 ± 82	1677 ± 337	1078 ± 272	677 ± 131	778 ± 113	9730
<i>AtRH40</i>	962 ± 55	634 ± 70	358 ± 65	379 ± 25	351 ± 32	688 ± 134	628 ± 33	289 ± 9	274 ± 21	4563
<i>AtRH42</i>	607 ± 67	355 ± 38	236 ± 30	205 ± 25	215 ± 21	467 ± 162	413 ± 69	181 ± 17	97 ± 9	2776
<i>AtRH56</i>	3164 ± 354	2148 ± 151	1119 ± 51	1078 ± 49	1161 ± 110	3825 ± 1810	2490 ± 557	1158 ± 127	1109 ± 154	17252
<i>AtRH58</i>	110 ± 6	496 ± 7	226 ± 20	87 ± 4	174 ± 17	239 ± 5	197 ± 55	93 ± 16	91 ± 10	1713
<i>AtEF1α-4</i>	84301 ± 16122	61696 ± 10722	31971 ± 5492	19875 ± 6353	29325 ± 5330	59120 ± 25284	58170 ± 13759	48501 ± 7951	22189 ± 7024	415148

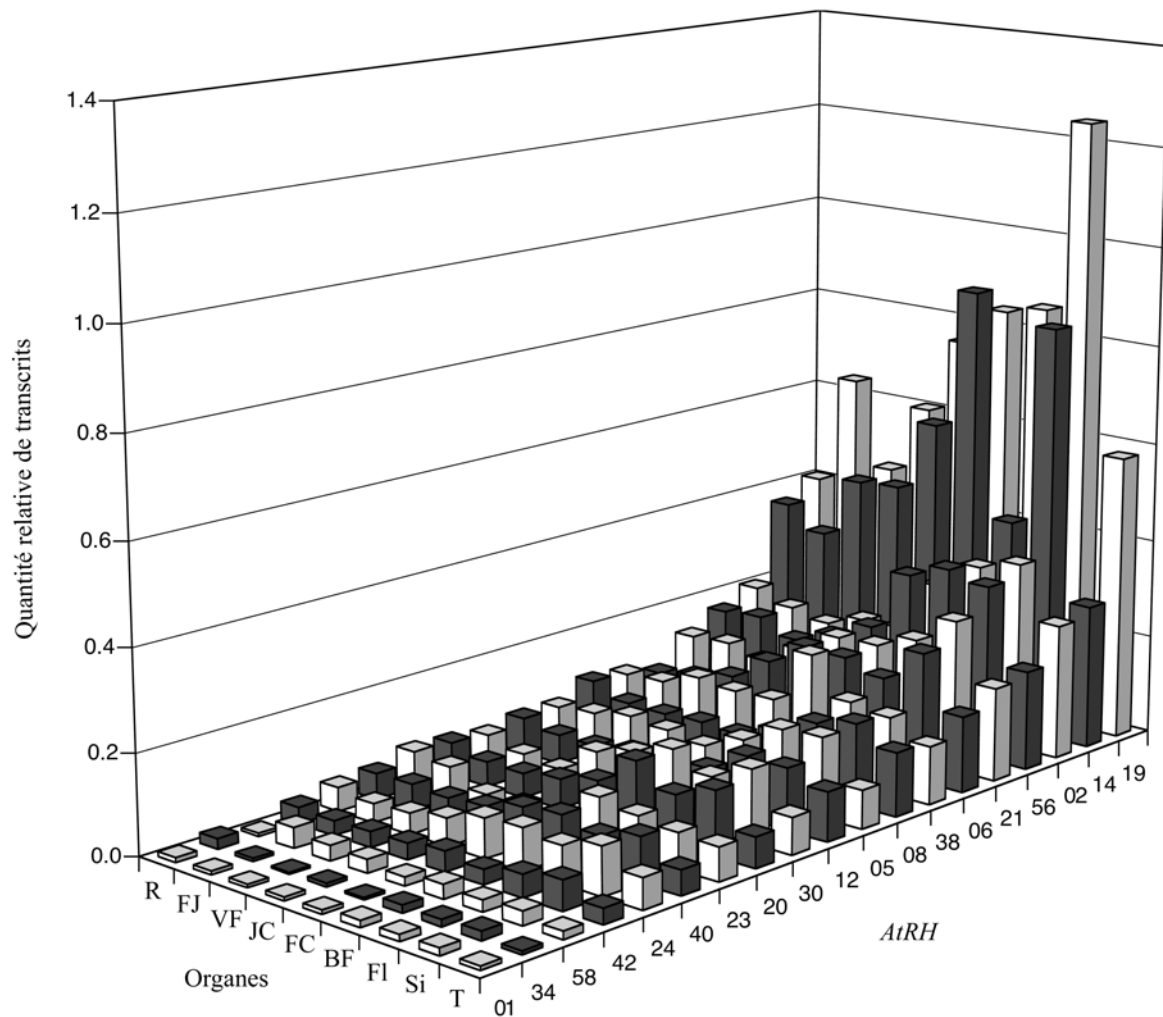
contrôle à chaque expérience, correspond à la moyenne des résultats de 21 à 23 expériences indépendantes, soit la moyenne de 42 à 46 valeurs. Dans quelques échantillons, pour les gènes très faiblement exprimés, seule la plus faible dilution a pu être utilisée. La valeur indiquée dans le tableau correspond alors à la moyenne de deux points. C'est le cas des gènes *AtRH1* (tous les organes), *AtRH40* (Si) et *AtRH58* (FJ, VF et BF).

L'intensité du signal obtenu correspond au bruit de fond de la PCR et est largement inférieure au niveau d'expression le plus bas mesuré parmi les *AtRH* (cf chapitre PCR quantitative, Tableau 8 p. 64). Aucune contamination n'est donc décelable dans ces échantillons.

Les 20 gènes *AtRH* étudiés sont exprimés dans tous les organes. Ils présentent des niveaux d'expression très différents. L'expression la plus faible est de 23 eg.ng<sup>-1</sup> ARN totaux pour *AtRH34* dans les feuilles caulines. L'expression la plus forte est de 18 679 eg.ng<sup>-1</sup> ARN totaux pour *AtRH4* dans les racines. En totalisant l'expression transcriptionnelle mesurée dans l'ensemble des organes pour chaque gène, *AtRH1*, gène le moins exprimé, apparaît 68 fois moins exprimé qu'*AtRH4*, gène le plus exprimé. De plus, le gène *AtRH4* présente la plus forte expression dans tous les organes à l'exception des fleurs. En effet, dans les fleurs, les gènes *AtRH4* et *AtRH19* présentent des niveaux de transcrits similaires. Les gènes *AtRH4* et *AtRH19* codent respectivement les facteurs d'initiation de la traduction, *eIF4A-1* et *eIF4A-2* (Metz *et al.*, 1992). Le gène *AtRH4* est en moyenne transcrit deux fois plus qu'*AtRH19*, et l'abondance de leurs transcrits diffère d'un facteur 1 dans les fleurs à un facteur 6 dans les racines.

Le gène *EF1 $\alpha$ -4* est souvent choisi comme témoin d'expression transcriptionnelle constitutive (Axelos *et al.*, 1989; Liboz *et al.*, 1990), aussi ai-je mesuré son expression par PCR quantitative en temps réel. Il existe une différence d'un facteur 4 entre son expression maximale dans les racines et minimale dans les feuilles caulines ce qui est cohérent avec une expression pseudo-constitutive. Pour comparaison, le gène *SUPI*, spécifiquement exprimé dans le pollen et les téguments de la graine, présente d'une part, une expression dans les fleurs et les siliques de 307 et 34 eg.ng<sup>-1</sup> ARN totaux respectivement et d'autre part, une expression nulle dans les racines, les feuilles ou les tiges (cf Tableau 8 p. 64).

En comparant l'expression des gènes *EF1 $\alpha$ -4* et *AtRH4* dans tous les organes, il apparaît que ces deux gènes présentent des profils similaires, le gène *EF1 $\alpha$ -4* étant cinq fois plus exprimé que le gène *AtRH4* en moyenne. Nous avons alors choisi le gène *AtRH4* pour normaliser l'expression des autres *AtRH* car son niveau d'expression est plus proche de celui des autres



**Figure 9** Abondance relative des transcrits de 20 *AtRH* par rapport à *AtRH4* dans différents organes d'*A. thaliana*.

L'abondance des transcrits de 20 *AtRH* mesurée par RT-PCR quantitative en temps réel (Tableau 9) a été rapportée à celle d'*AtRH4*, gène le plus exprimé de la famille. Les 20 gènes *AtRH*, indiqués par leur numéro, sont ordonnés selon leur niveau global d'expression. Les organes analysés proviennent du même lot de 96 plantes ayant d'abord poussé en jours courts jusqu'au stade rosette 8 à 12 feuilles. Les racines (R), les jeunes feuilles de la rosette (FJ) et les feuilles âgées de la rosette (VF) ont été prélevées à ce stade sur 48 plantes. Puis, les 48 plantes restantes ont été transférées en jours longs. Les jeunes feuilles autour des bourgeons (JC), les feuilles caulinaires matures (FC), les tiges des hampes florales (T), les bourgeons floraux (BF), les fleurs (FI) et les siliques (Si) ont été récoltés un mois après la transition florale.

*AtRH*. En effet, plus l'écart est grand entre la valeur de référence et la valeur à normaliser, plus l'erreur introduite par une faible variation est amplifiée.

La Figure 9 représente le niveau d'expression normalisé des 20 *AtRH* dans les différents organes. Les gènes *AtRH19* et *AtRH14* présentent une forte activité transcriptionnelle proche du niveau d'*AtRH4*. Ensuite viennent les gènes *AtRH2*, *AtRH56*, *AtRH21*, *AtRH6* et *AtRH8* dont le niveau d'expression moyen est compris entre 4 et 6 fois moins que celui du gène *AtRH4*. Puis, on trouve, dans un ordre d'expression décroissante, les gènes *AtRH38*, *AtRH12*, *AtRH5*, *AtRH30*, *AtRH23*, *AtRH24*, *AtRH20* et *AtRH40*. Enfin quatre gènes sont très faiblement exprimés (20 à 100 fois moins exprimés que le gène *AtRH4*). Il s'agit des gènes *AtRH42*, *AtRH58*, *AtRH34* et *AtRH1*.

L'expression des 21 *AtRH* présentés dans ce manuscrit couvre donc une large gamme de niveaux. De plus, ces différents niveaux se répartissent selon une échelle continue plutôt que par groupes bien séparés.

## 2. Northern blots

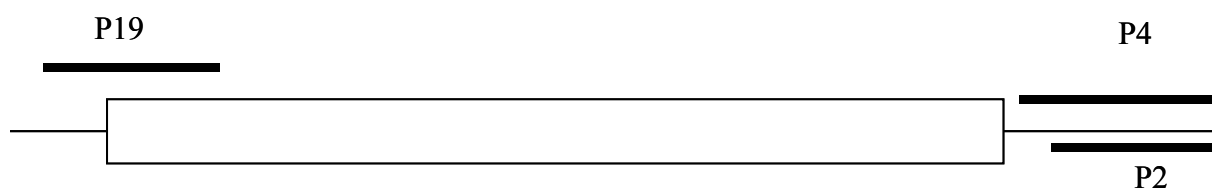
Nous avons déterminé le niveau d'expression de plusieurs *AtRH* dans les jeunes feuilles de la rosette par une méthode différente de la PCR quantitative en temps réel. Il nous semblait intéressant de choisir des paralogues suffisamment exprimés et suffisamment proches en séquence afin de pouvoir les détecter par hybridation et ainsi comparer la spécificité de cette méthode avec celle de la PCR quantitative. Les trois gènes *AtRH2*, *AtRH4* et *AtRH19* ont été retenus. En effet, d'après le nombre d'EST disponibles dans les bases de données, ils présentent un niveau d'expression transcriptionnel suffisant pour être détectés par hybridation<sup>12</sup>. De plus, les gènes *AtRH4* et *AtRH19* sont issus de la duplication du même gène ancestral. Leurs ARNm présentent 80% de nucléotides identiques. Le gène *AtRH2* n'appartient pas au même groupe fonctionnel qu'*AtRH4* et *AtRH19*. Cependant, c'est le gène le plus exprimé le plus proche en séquence d'*AtRH4* et *19*. En effet, le groupe fonctionnel auquel il appartient possède un ancêtre commun avec celui d'*AtRH4* et *AtRH19*. La séquence de son ARNm présente respectivement 58% et 55% de nucléotides identiques à celles d'*AtRH4* et *AtRH19*. L'alignement de la séquence des ARNm de ces trois gènes a permis de placer les sondes P2, P4 et P19 au niveau des parties les plus divergentes. Les sondes P2 et P4

---

<sup>12</sup> La relation entre le nombre d'EST et la teneur en transcrite est abordée dans le point suivant (3. Northern in silico) et présentée dans la figure 10c.

**Figure 10 Position des sondes par rapport aux régions codante et non codantes des transcrits *AtRH2*, *AtRH4* et *AtRH19*.**

La séquence codante de l'ATG au stop est représenté par une boîte. Les traits symbolisent les régions 5' et 3' non traduites. La position des sondes spécifiques des transcrits *AtRH2*, *AtRH4* et *AtRH19* est indiquée par les traits épais associés aux sigles P2, P4 et P19 respectivement.



**Tableau 10 Caractéristiques des sondes utilisées pour l'hybridation.**

Une ligne indique successivement le nom du gène d'intérêt, la séquence 5'-3' des amorces utilisées pour l'amplification de la sonde, la longueur en nucléotides, le pourcentage en GC ainsi que le pourcentage d'identité avec les différents transcrits *AtRH*.

Gène	Amorce sens	Amorce antisens	nt	%GC	P2	P4	P19
<b>AtRH2</b>	TAACGGTGTATTTCTTTGATG	AACCGTTTCAACTGTCGAGC	225	40	100	40	23
<b>AtRH4</b>	CTGTTGCTGTGTTTGTCTG	ATCAAAGACAAACAACAAAGCC	260	37	43	100	48
<b>AtRH19</b>	CACCAGATCCAAGTTCGC	ATCAAAGCTCTCTCATGGACCTC	227	43	41	72	100
<b>ACT2/8</b>	GTTTTTCCCAGTGTGTTG	GTCATCTTCTCTCTGTTGGC	274	45	-	-	-

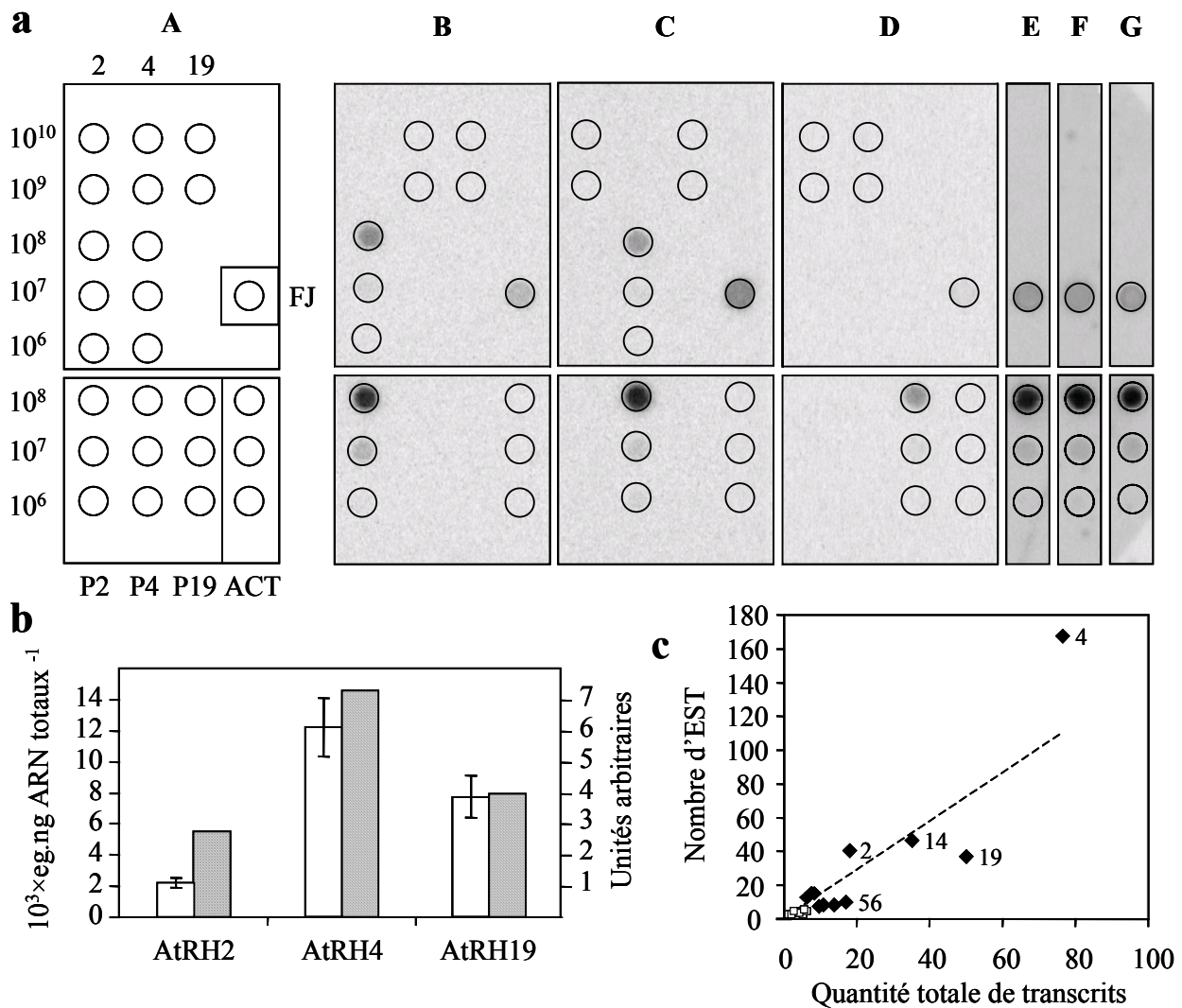


ont été placées à l'extrémité 3' non codante des gènes *AtRH2* et *AtRH4* respectivement. La sonde P19 correspond à une région de l'extrémité 5' du gène *AtRH19* chevauchant les parties codante et non codante (Figure 10). Les gènes d'actine appartiennent à une famille multigénique chez *A. thaliana* et leur expression transcriptionnelle varie beaucoup d'un membre à l'autre (Huang *et al.*, 1997). Cependant, il a été montré que les gènes *ACT2* et *ACT8* présentent des profils d'expression complémentaires (An *et al.*, 1996b). C'est pourquoi la sonde ACT a été positionnée dans une partie commune des transcrits *ACT2* et *ACT8*. L'homogénéité des sondes a été recherchée en sélectionnant des régions de taille (225 à 274 nt) et de composition (37 à 45% GC) similaires. Les pourcentages d'identité des sondes avec les différents transcrits varient de 72% à 23% (Tableau 10).

Les ADNc clonés des gènes *AtRH2*, *AtRH4* et *AtRH19* ont été utilisés sur la première partie des membranes comme contrôle des hybridations croisées d'une part ( $10^9$  et  $10^{10}$  copies) et comme gamme-étalon d'autre part ( $10^6$ ,  $10^7$  et  $10^8$  copies). Les ADNc des gènes *AtRH2* et *AtRH4* sont pleine longueur. Par contre, l'ADNc dont nous disposons pour le gène *AtRH19* est tronqué en 5'. Il ne contient pas la partie correspondant à la sonde P19. Par contre, son extrémité 3' présente une forte ressemblance avec *AtRH4* : 72% des nucléotides sont identiques. Nous avons donc utilisé cet ADNc tronqué comme contrôle d'hybridation croisée des sondes P2 et P4 sur *AtRH19*. Sur cette même partie a été déposée une quantité d'ADNc simple brin équivalant à 20 µg d'ARN totaux de jeunes feuilles de rosette (Figure 11).

Pour mesurer de façon homogène l'abondance des transcrits *AtRH* et *ACT*, des gammes de produits PCR non marqués radioactivement ( $10^6$ ,  $10^7$  et  $10^8$  copies) correspondant aux quatre sondes P2, P4, P19 et ACT ont été déposées sur la deuxième partie des membranes. Les trois membranes ont été hybridées avec l'une des trois sondes P2, P4 ou P19. Après déshybridation, les trois membranes ont été ré-hybridées avec la sonde ACT.

Le schéma global des dépôts (A) ainsi que les images des différentes hybridations (B à G) sont regroupées dans la Figure 11a. Les cercles noirs sur les images d'hybridation symbolisent les dépôts effectifs réalisés sur les différentes membranes (B à G). Dans nos conditions d'hybridation et de lavage, nous n'observons aucun signal non spécifique ni hybridation croisée. Les quatre sondes P2, P4, P19 et ACT sont donc spécifiques. Ainsi, malgré la proximité des séquences, il est possible de distinguer les trois paralogues *AtRH2*, *AtRH4* et *AtRH19* par hybridation.



**Figure 11 Comparaisons de la mesure des transcrits par Northern blots, Northern in silico et RT-PCR.**

**a.** Estimation de la quantité de transcrits d'*AtRH2*, *AtRH4*, *AtRH19* et *ACT2/ACT8* par dot blots dans les jeunes feuilles de la rosette d'*A. thaliana*.

A: diagramme de dépôt des échantillons. Le nombre de copies est indiqué sur la gauche. Pistes 2 et 4 : plasmides contenant l'ADNc pleine longueur d'*AtRH2* et *AtRH4*, respectivement. Piste 19 : plasmide contenant un ADNc partiel d'*AtRH19*, tronqué en 5'. P2, P4, P19 et ACT : sondes froides (produits PCR) spécifiques d'*AtRH2*, *AtRH4*, *AtRH19* et *ACT2/ACT8* respectivement. FJ: ADNc simple brin obtenu à partir de 20µg d'ARN totaux de jeunes feuilles de la rosette d'*A. thaliana*. B, C, D : Le diagramme de dépôt effectif des échantillons est superposé aux signaux d'hybridation des sondes spécifiques P2 (B), P4 (C) et P19 (D). E, F, G: Signaux des 3 membranes précédentes (B, C, D) hybridées avec la sonde ACT. Seule la dernière colonne est montrée.

**b.** Quantité de transcrits d'*AtRH2*, *AtRH4* et *AtRH19* mesurée par RT-PCR (en blanc) et par Northern blots (en gris) dans les jeunes feuilles de la rosette.

Les barres verticales représentent l'écart-type de la moyenne des valeurs obtenues par RT-PCR quantitative en temps réel. Les quantités de transcrits observées en Northern blot ont été estimées en analysant au phosphorimager® l'intensité des signaux d'hybridation. L'intensité des signaux d'hybridation des sondes spécifiques a été rapportée à celle de l'actine.

**c.** Nombre d'EST et quantité totale de transcrits estimée par RT-PCR de 21 *AtRH*.

Les carrés blancs représentent gènes ayant moins de 9 EST associées et les losanges noirs représentent les gènes ayant plus de 8 EST associées. Seuls sont indiqués les numéros des gènes les plus exprimés (expression totale supérieure à 15.10<sup>3</sup> eg.ng<sup>-1</sup> ARN totaux). La quantité totale de transcrits mesurée par PCR quantitative en temps réel (Tableau 9) est exprimée en 10<sup>3</sup> eg.ng<sup>-1</sup> ARN totaux. Le nombre d'EST provient de l'analyse de la version 080103 de la base de données dbEST du NCBI. Si plusieurs EST sont associées au même clone ADNc (typiquement l'une marquant l'extrémité 5' et l'autre l'extrémité 3'), une seule EST a été comptée.

L'abondance des transcrits des gènes *AtRH2*, *AtRH4*, *AtRH19* (B, C et D respectivement) et *ACT2/ACT8* (E, F, G) a été estimée par *dot blot* dans des ADNc simple brin de jeunes feuilles de la rosette. L'image numérique obtenue au PhosphorImager® a été analysée à l'aide du logiciel ImageQuant®. L'intensité du signal due à l'hybridation des sondes P2, P4, P19 et ACT sur l'équivalent ADNc simple brin de 20 µg d'ARN totaux de jeunes feuilles de la rosette a été comparée avec celle des gammes de sondes froides ( $10^6$ ,  $10^7$  et  $10^8$  copies). L'intensité du signal observé pour l'actine est similaire sur les trois membranes (Figure 11a). Par contre, l'intensité du signal observé pour chaque *AtRH* est différente. Après standardisation de l'intensité des signaux avec l'actine, l'abondance des transcrits des gènes *AtRH2*, *AtRH4* et *AtRH19* dans les jeunes feuilles de la rosette obtenue par *Northern blot* a été comparée aux valeurs obtenues dans le même extrait par PCR quantitative (Figure 11b).

Les résultats des deux méthodes concordent : dans les deux cas, *AtRH4* est 1,5 à 2 fois plus exprimé qu'*AtRH19* et 3 à 5,5 fois plus qu'*AtRH2*. Les deux techniques sont donc valables pour étudier l'abondance des transcrits des gènes d'une famille multigénique. Cependant la technique de *dot blot* est plus coûteuse en temps et en matériel que la PCR quantitative. Elle est également moins sensible et moins adaptée pour les gènes faiblement exprimés.

### 3. *Northern in silico*

#### a. Expression et nombre d'EST

Le principe du « *Northern in silico* » consiste à déterminer le nombre d'EST d'un gène dans les bases de données pour caractériser son expression transcriptionnelle. Cette méthode est basée sur l'hypothèse que plus un gène est exprimé, plus les EST disponibles pour ce gène sont abondantes. Les EST correspondant à chaque *AtRH* ont été recherchées dans la base de données dbEST (version 080103). Compte-tenu des incertitudes provenant de séquences incomplètes ou de mauvaise qualité, l'attribution de chaque EST à un seul gène a été manuellement vérifiée avant de calculer le nombre d'EST pour chaque gène. Parallèlement, le niveau total d'expression transcriptionnelle de chaque *AtRH* a été estimé en additionnant l'abondance des transcrits dans chaque organe mesurée par PCR quantitative (Tableau 9). Ensuite, le nombre de clones disponibles dans la base de données dbEST pour chaque gène *AtRH* a été représenté en fonction de son niveau d'expression totale mesuré par PCR quantitative (Figure 11c).

Les *AtRH* peuvent être répartis en quatre groupes selon leur expression totale et leur nombre d'EST. Le premier groupe correspond aux gènes présentant une expression totale inférieure ou égale à  $6.10^3$  eg.ng<sup>-1</sup> ARN totaux. Il s'agit des gènes *AtRH1*, 20, 23, 24, 30, 34, 40, 42 et 58 (carrés blancs, Figure 11c). Ils sont associés à 6 EST au maximum. Les gènes pour lesquels au moins 9 EST sont disponibles (représentés en noir) se répartissent dans les trois autres groupes. Les gènes *AtRH5*, 6, 8, 12, 21, 38 et 56 présentent une expression totale comprise entre  $6,2.10^3$  et  $17,3.10^3$  eg.ng<sup>-1</sup> ARN totaux et 9 à 17 EST associées. Ils forment donc le second groupe. Le troisième groupe est composé des gènes *AtRH2*, 14 et 19. Ils présentent une expression totale comprise entre  $18,2.10^3$  et  $50,1.10^3$  eg.ng<sup>-1</sup> ARN totaux et un nombre d'EST associées compris entre 39 et 47. Le gène *AtRH4* forme le dernier groupe à lui seul. Il présente une expression totale de  $76,5.10^3$  eg.ng<sup>-1</sup> ARN totaux et 169 EST correspondant à ce gène sont disponibles dans la base de données dbEST.

Il existe une corrélation positive ( $r^2 = 0,98$ ) entre le nombre d'EST et le niveau d'expression mesuré par PCR quantitative des trois groupes de gènes associés à plus de 8 EST. Globalement, l'hypothèse de départ est vérifiée : plus le niveau d'expression d'un groupe de gènes est fort, plus le nombre d'EST associé à ce groupe est élevé. Par contre, cette relation n'est pas valable pour un gène isolé. En effet, il peut y avoir une différence d'un facteur 5 entre l'expression totale de gènes possédant le même nombre d'EST. Pour certains gènes, un nombre d'EST élevé est associé à une expression faible.

Le nombre d'EST peut être utilisé comme un indicateur grossier du niveau d'expression d'un gène *AtRH* : moins de 10 EST, expression faible ; de 10 à 100 EST expression moyenne ; plus de 100 EST expression forte. Cependant, plusieurs faits viennent limiter la valeur de ces chiffres.

Tout d'abord, le nombre d'EST est associé à la version de la base de données dbEST. Les chiffres avancés ici sont donc valables pour la version du 08 janvier 2003. Comme le nombre de séquences déposées augmente régulièrement, on s'attend à ce que ces chiffres augmentent avec le temps. Cependant, il est difficile de prévoir leur évolution. En effet, si la base de donnée dbEST est enrichie de façon homogène, le nombre d'EST correspondant à des gènes faiblement exprimés augmentera proportionnellement à celui des gènes fortement exprimés. Mais ce n'est en général pas le cas. La nature des séquences déposées dépend des projets de séquençage d'EST. Le contexte de génération des banques d'ADNc, tout comme un choix concernant la taille des fragments clonés, influencera la proportion d'EST d'une catégorie de

gènes par rapport à une autre. Typiquement, selon que les transcrits sont extraits de plantes entières ou d'un organe particulier comme les racines ou les fleurs, la population d'EST déposée sera différente. Pour des gènes exprimés dans tous les organes végétaux, tels que les *AtRH*, ceci n'introduira pas trop de variations du moment que les EST issues de banques dites « spécifiques » restent dans des proportions proches. Cependant, la diversification de l'origine des banques entraînera une augmentation de la variation du nombre d'EST disponibles pour chaque gène, rendant de moins en moins possible l'établissement d'une relation entre nombre d'EST et expression transcriptionnelle.

De plus, un faible nombre d'EST n'est pas forcément synonyme d'expression faible, notamment pour un gène exprimé de façon spécifique. En effet, l'expression d'un gène présentant 9 EST, provenant toutes uniquement d'une banque d'ADNc de racines, sera considérée comme forte dans cet organe, alors que ce chiffre représente une expression faible pour un gène *AtRH*.

L'augmentation du nombre d'EST dans les bases de données a une autre conséquence : celle de rendre de plus en plus fastidieuse la recherche exhaustive des EST associées à chaque gène et l'établissement de la relation univoque liant chaque EST à un gène particulier de la famille. En effet au sein d'une famille, les gènes présentent des séquences proches. Certains gènes présentent des séquences nucléiques identiques à plus de 95%. Or, les séquences d'EST sont issues d'un séquençage unique de l'extrémité 5' ou 3' d'un ADNc et présentent jusqu'à 5% d'erreurs. Pour comparaison, les séquences génomiques sont en principe séquencées 6 fois et présentent moins de 0,5% d'erreurs. C'est pourquoi il peut être difficile de déterminer de quel gène proviennent certaines EST. De plus, dans le résultat d'une recherche par le logiciel BLAST, les séquences d'EST proposées correspondent souvent à plusieurs gènes et arrivent mélangées. Une expertise humaine est donc nécessaire pour ne sélectionner que les EST du gène d'intérêt. Lorsqu'il s'agit d'expertiser plusieurs dizaines de séquences voire plusieurs centaines pour un gène donné (dans la version 080103 de dbEST, plus de 400 EST étaient associées au gène *AtEF1 $\alpha$ -4* ), le risque d'erreur n'est donc pas négligeable. Une automatisation de cette expertise humaine est possible mais elle nécessite de disposer soit de l'ensemble des gènes de la famille, soit de l'ensemble du génome, comme ceci a été réalisé dans la base de données FLAGdb<sup>++13</sup>. FLAGdb<sup>++</sup> est une base de données intégrative développée autour du génome d'*A. thaliana*. Pour associer automatiquement les EST et les

---

<sup>13</sup> <http://genoplante-info.infobiogen.fr/FLAGdb>

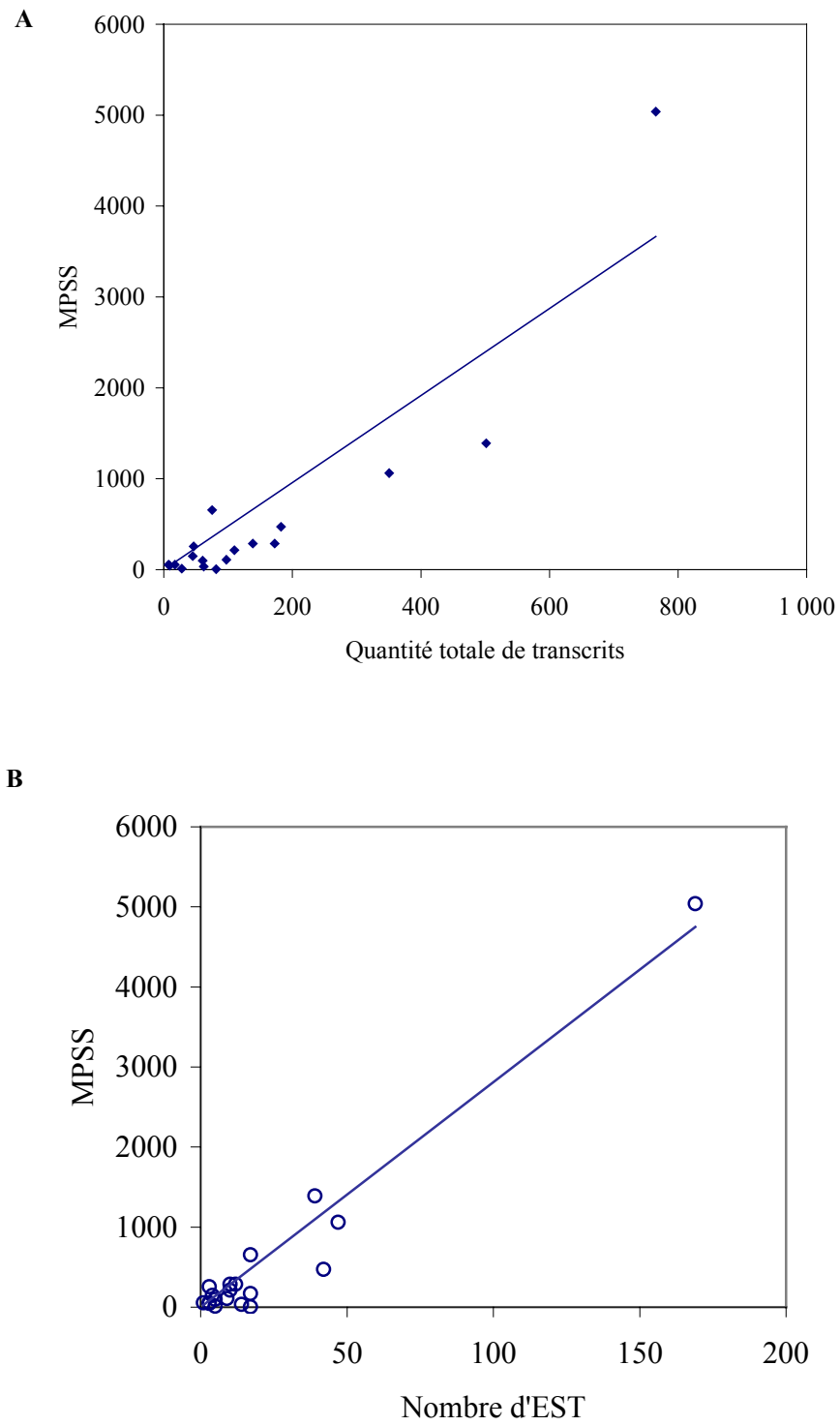


Figure 12 Comparaison des données MPSS avec l'expression totale mesurée par RT-PCR quantitative en temps réel ( $R^2=0,81$ ) et avec le nombre d'EST ( $R^2=0,94$ ).

ADNc à chaque gène, l'ensemble des séquences des transcrits de la base de données GenBank (mars 2003) a été aligné aux chromosomes d'*A. thaliana* à l'aide du logiciel ACEMBLY. Ce logiciel permet de reconstruire de longues séquences comme des chromosomes à partir de l'alignement de séquences d'origine multiple. Dans FLAGdb<sup>++</sup>, le nombre d'EST et d'ADNc associés à chaque gène a été choisi comme critère pour représenter ces données. Les trois classes retenues – de 1 à 10 EST et ADNc, de 11 à 50 EST et ADNc et plus de 50 EST et ADNc<sup>14</sup> – correspondent à mes estimations sur la famille des *AtRH*.

#### b. Expression et MPSS

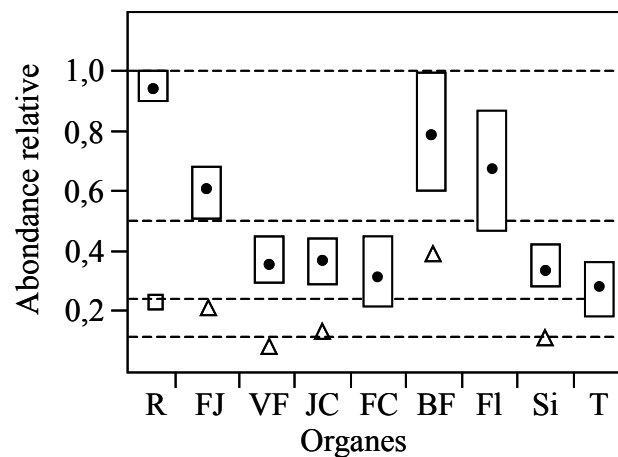
L'informatique est un outil puissant et pratique pour la manipulation des séquences et pour effectuer des calculs sur leur fréquence. Des sites permettant d'estimer l'expression transcriptionnelle d'un gène par « *Northern in silico* » à partir d'EST ou d'étiquettes SAGE (*Serial Analysis of Gene Expression*) par quelques clics de souris ont vu le jour<sup>15</sup>. Une nouvelle technique, appelée MPSS pour *Massively Parallel Signature Sequencing* (séquençage massif de signatures en parallèle), a été inventée et commercialisée par Lynx Therapeutics (Californie). Elle est basée sur un principe similaire aux étiquettes SAGE. Cette technique permet de générer des étiquettes à partir de positions connues sur les ARNm (Brenner *et al.*, 2000a; Brenner *et al.*, 2000b). L'hypothèse proposée est que l'abondance relative de ces étiquettes MPSS donne une estimation de l'expression des gènes. Les étiquettes MPSS font 17 nucléotides de long, et permettent d'identifier plus de 95% des gènes d'*A. thaliana*. Des banques d'étiquettes MPSS ont été réalisées sur des calcs de cellules, des fleurs, des racines, des organes aériens et des siliques d'*A. thaliana*. Via l'interface d'interrogation de la base de données<sup>16</sup>, il est possible de connaître les étiquettes associées à un gène donné ou d'effectuer des requêtes plus complexe du type « Quelles sont les étiquettes MPSS 10 fois plus fréquentes dans les fleurs que dans les racines ? ».

J'ai relevé l'abondance relative des étiquettes MPSS associées aux 21 *AtRH* de mon étude et j'ai comparé ces données aux résultats de PCR quantitative en temps réel et au nombre d'EST (Figure 12). Il existe une meilleure concordance entre les estimations de l'expression par MPSS et EST que par MPSS et PCR quantitative. Ceci provient probablement d'un même biais technique des protocoles générant les EST et les MPSS. En effet, ces deux méthodes sont basées sur le clonage des ARNm polyadénylés, contrairement à la PCR quantitative.

<sup>14</sup> <http://genoplante-info.infobiogen.fr/FLAGdb/ OnLineHelp/EST.html>

<sup>15</sup> <http://www.ncbi.nlm.nih.gov/geo/>

<sup>16</sup> <http://dbixs001.dbi.udel.edu/MPSS4/java.html> et <http://mpss.udel.edu/at/>



**Figure 13** Profils d'expression transcriptionnelle de 21 *AtRH* dans différents organes d'*A. thaliana*.

L'expression transcriptionnelle de chaque *AtRH* mesurée par RT-PCR quantitative en temps réel dans chaque organe (Tableau 9) a été rapportée à son expression dans l'organe pour lequel elle est la plus forte. Les profils individuels ainsi obtenus ont été superposés et comparés. Les rectangles blancs contiennent  $\frac{2}{3}$  des valeurs réparties autour de la moyenne (points noirs). Les valeurs restantes appartenant à l'intervalle de confiance contenant 95% des valeurs ( $x \pm 1,96\sigma$ ) ne sont pas représentées. Les valeurs en dehors de l'intervalle de confiance et supérieures à 2 fois ou inférieures à 0,5 fois la moyenne sont représentées par des symboles. Cette dernière catégorie comprend des valeurs de deux gènes : elles sont indiquées par des triangles pour *AtRH34* et par un carré pour *AtRH58*. R, racines; FJ, jeunes feuilles de la rosette; VF, feuilles âgées de la rosette; JC, jeunes feuilles caulines; FC, feuilles caulines matures; BF, bourgeons floraux; FI, fleurs; Si, siliques et T, tiges des hampes florales.



La PCR quantitative a permis de mesurer d'expression de 23 gènes sur une large gamme de valeurs s'étendant de 23 eg.ng<sup>-1</sup> ARN totaux à 84301 eg.ng<sup>-1</sup> ARN totaux. Par ailleurs, il semble qu'en dessous de 10 EST et 200 MPSS, il n'y ait pas de relation entre le nombre d'étiquettes et l'expression transcriptionnelle. En considérant ces limites, l'expression de 8 et 9 gènes peut être estimée par le nombre d'EST ou de MPSS, respectivement. La PCR quantitative est donc la méthode d'estimation de l'expression transcriptionnelle la plus sensible.

### C. Profils d'expression transcriptionnel

L'existence de profils d'expression spécifiques, évitant la redondance fonctionnelle, est l'une des explications donnée du maintien d'un grand nombre de paralogues dans les génomes (Force *et al.*, 1999; Lynch & Force, 2000; Lynch *et al.*, 2001). Pour rechercher l'existence de tels profils, l'expression transcriptionnelle de chaque gène mesurée par PCR quantitative dans chaque organe a été rapportée à son expression dans l'organe pour lequel elle est maximale.

Tout d'abord, 19 gènes *AtRH* parmi les 21 étudiés présentent une différence d'un facteur 3 à 6 entre leur expression minimale et maximale, avec une moyenne de 4,2. Ceci est cohérent avec une expression pseudo-constitutive attendue pour des gènes intervenant dans le métabolisme de l'ARN. Les deux gènes restants, *AtRH23* et *AtRH34*, présentent une différence d'un facteur 17 entre leur expression minimale dans les feuilles caulines et leur expression maximale dans les racines.

Ensuite, tous les gènes, à l'exception d'*AtRH34* et *AtRH58*, suivent un profil similaire, indépendamment de leur niveau d'expression (Figure 13). Ce profil consiste en une expression transcriptionnelle maximale dans les racines, puis une expression de 95 à 50% de l'expression dans les racines dans les boutons floraux, les fleurs et les jeunes feuilles de la rosette, et enfin, une expression entre 50 et 25 % de l'expression dans les racines dans les autres organes testés, c'est-à-dire les feuilles âgées de la rosette, les jeunes feuilles caulines autour des boutons, les feuilles caulines, les tiges et les jeunes siliques. Les organes dans lesquels l'activité transcriptionnelle des *AtRH* est la plus forte correspondent aux organes dans lesquels on trouve le plus de cellules actives, ce qui est cohérent avec leur implication dans le métabolisme des ARN. Le profil d'expression de ces *AtRH* pourrait refléter le niveau d'activité transcriptionnelle globale des cellules d'*A. thaliana*.

Quelques variations à ce profil global existent. Alors que 15 gènes sur 20 présentent une expression maximale dans les racines, l'expression des gènes *AtRH5*, *AtRH6* et *AtRH56* est maximale dans les bourgeons floraux, celle d'*AtRH19* est maximale dans les fleurs et celle

d'*AtRH58* dans les jeunes feuilles de la rosette. Toutefois, excepté pour *AtRH58*, le profil global est respecté car l'expression de ces gènes dans les racines reste forte, comprise entre 75% et 86% de l'expression maximale.

Parmi les deux gènes montrant un profil d'expression différent, *AtRH34* et *AtRH58* (Figure 13), le gène *AtRH58* présente une activité transcriptionnelle minimale dans les racines (22%) et maximale dans les jeunes feuilles de la rosette. Ce profil spécifique concorde avec le fait que, sur la base de la séquence protéique, la protéine AtRH58 serait adressée au chloroplaste selon la prédiction du logiciel PREDOTAR 1.03<sup>17</sup> (score = 0,89). De plus, alors qu'il n'a pas été identifié d'orthologue au gène *AtRH58* parmi des gènes d'hélicases à boîte DEAD de *S. cerevisiae*, de *C. elegans* ni de *D. melanogaster*, l'inactivation d'un orthologue potentiel d'*AtRH58* chez le Tabac entraîne de grands bouleversements dans la biogenèse et la différenciation des plastes (Wang *et al.*, 2000b). Enfin, les 500 pb en amont du site d'initiation de la traduction contiennent des éléments similaires à ceux qui sont impliqués dans la réponse à la lumière d'autres gènes de plantes (Arguello-Astorga & Herrera-Estrella, 1996; Park *et al.*, 1996; Feldbrugge *et al.*, 1997). Le gène *AtRH58* présente donc un profil d'expression transcriptionnel spécifique et des éléments structuraux qui concourent à l'impliquer dans une fonction photosynthétique.

Le gène *AtRH34* présente un niveau d'expression relatif beaucoup plus bas dans les feuilles de la rosette, qu'elles soient jeunes ou âgées, les jeunes feuilles caulines, les bourgeons floraux et les siliques (Figure 13). Dans les autres organes, ce gène présente une activité transcriptionnelle conforme au profil global de la famille. Les gènes *AtRH2* et *AtRH34* appartiennent au groupe d'orthologues potentiels du gène *ScRH-FAL1*. Les *AtRH* du groupe FAL1 ont un ancêtre commun avec les *AtRH* du groupe *TIF1/TIF2* mais présentent une structure intron-exon complètement différente (Boudet *et al.*, 2001). Selon l'organe, *AtRH34* est 10 à 50 fois moins transcrit qu'*AtRH2*. Cette grande différence d'expression entre les organes, non observée pour les autres *AtRH*, provient du fait qu'*AtRH34* est relativement moins transcrit dans les tissus foliaires. Le profil d'expression transcriptionnel d'*AtRH34* suggère que le produit de ce gène puisse être associé à une fonction spécifique des organes puits comme les racines ou les fleurs et moins exprimé dans les organes sources. Par ailleurs, les protéines AtRH2 et AtRH34 présentent des points isoélectriques différents : 5,9 et 8,2

---

<sup>17</sup> <http://genoplante-info.infobiogen.fr/predotar>

respectivement, suggérant une fonction dans des environnements cellulaires différents. Cet élément est en faveur d'une divergence d'*AtRH34* vers une nouvelle fonction après la duplication de *AtRH34* et *AtRH2* chez les plantes.

Ainsi, les deux cas de transcription spécifique de gènes de la famille *AtRH*, à savoir les gènes *AtRH58* et *AtRH34*, sont associés respectivement à deux fonctions majeures spécifiques des plantes, à savoir la photosynthèse et l'import des photoassimilats.

Une approche de complémentation des mutants de levure *FAL1* et *TIF1/TIF2* par les gènes *AtRH2*, *AtRH4*, *AtRH19*, *AtRH23* et *AtRH34* a été entreprise en collaboration avec M-C Daugeron pour tester la conservation de fonction entre hélicases de levure et de plante. Malheureusement, les clones des gènes *AtRH* obtenus présentaient des erreurs de nucléotides, modifiant des acides aminés conservés du site catalytique des hélicases à boîte DEAD. De ce fait, les tests de complémentation n'ont pas été réalisés, faute d'avoir obtenu des clones sans erreur.

#### D. Conclusion

Dans cette partie, nous avons montré l'existence, d'une part, d'un profil d'expression caractéristique, et d'autre part, d'une large gamme de niveaux d'expression. La conservation d'un profil d'expression caractéristique des *AtRH* suggère l'existence d'un mécanisme de régulation commun à l'ensemble des gènes suivant ce profil et donc la conservation d'éléments régulateurs au sein de ces gènes. La diversité des niveaux d'expression implique un contrôle strict de l'abondance des transcrits des *AtRH*. Ce contraste entre la conservation des profils et la diversité des niveaux d'expression indique un contrôle fort de l'expression de cette famille de gènes, ce qui sous-entend l'existence de signaux de régulation précis. Afin d'identifier des éléments potentiellement impliqués dans le contrôle de l'expression des *AtRH*, nous avons analysé la séquence promotrice des 56 gènes *AtRH*.

*AtRH01* AATCGGTTTCATGGTTAACAATTAACAAGTAAATGTAAAACCCATTGGTACACCCCTTGCC  
*AtRH02* GATTTACAATCGTACCCTATTTTTAAGCCGCGCTTATCTCGAACCTATACCTGTAGAGA  
*AtRH03* GATATTTTACATCAACTTGAGTGGTAGA **AATAAATA** AAGATAGGGCAGAAGCAGAGAAGC  
***AtRH04*** TCCAAAGGAGAAAACCCTAAAGACGGAG **TATATAAA** CAAGGTAACGCGTTTTCTCTCAGC  
*AtRH05* TGCAATTGTAGAAAATATTTTTGCAATGC **TTTAAA** ACCCTAAAAACCTAAATTGAACAAC  
*AtRH06* ATTGTTAGCATATTCGGATCTTGCCCTCCCCCGAGAGAGAGAGTCCAATCAATCAATCAA  
*AtRH07* TTAGGGTTTTCTCTAAAATCTCAAAAC **TATATAAA** AAGAGAGGTAAACCCCTAGCAATTGTC  
*AtRH08* CTAACGAAATCCGACGAAATCACAAACAT **TATAAT** ATTACATAGAGATAATCATCTTTCT  
*AtRH09* GCAGGTGATTATAACAAAGTTCCCTTTTATACCCGCTTTTGATTCTTCTTTAAATCTCT  
*AtRH10* CCAATTAGGT **TATTAA** AATAACAACATAACTAAAACCCCTTGATAGCTAGAACTCGCTCTC  
***AtRH11*** TAATTAGATTTTAAATCAATTCTTAAA **TATAAA** ACAAGCTAATCTCATTATCATACCAT  
*AtRH12* ATTTAACGTAAACATAAAAAGGCCATTTTGA **TAAATA** GTGAGAGTCTGAAAAAAGAAAGA  
*AtRH13* TATACTACATGGTTTCGTTCTGATTTAGGGTTTATGCAAGTTTCAACTTCTCTTGAGAT  
***AtRH14*** TGCAGATATTTTTGTTGCGACGATGGATGGTTTT **TATAAA** CCTAACTAACAGAGACGGTCT  
***AtRH15*** TTGTCAACCCTATCCAGTCTCCTTG **TATATA** TTTATTTACGACACCAACGCGGCTTGGT  
*AtRH16* GCCCAATATTGGAGTCATCAATCAGGTCTGGGACATTGAAAAACTCAGTAACAGTCTAA  
*AtRH17* no 5' UTR  
*AtRH18* CAGTCATTGCAAGGTTACTGCAAGGTTAAAGCGGGAAGTGAGCAACGAGAGGCAGCGGCG  
***AtRH19*** CCTAAAAAAGAGGGGGAC **TATATA** ACTGGCTAACACGTATTGGGTTGAGTCTCTC  
*AtRH20* GGCACCTCTCACCTTTTCGTCGTCCTCAATTTGTAACCCCTAACGACTAGATTTTCAGTACCT  
*AtRH21* GCTAGAATTGGGGGGTTTTGGAAAAATGTAGAAAAAGAGAATCCAAGTGAAAAAGAAATC  
*AtRH22* GAAGAGAGAGCTGAAACAGAGGTTTCTTACTTTCTTCTCTGTATCTCTCATATTTTGCTT  
***AtRH23*** GTTTAGAGGGCAAATCGGGAAGA **TATTTA** AGTTGTTTTCTTTTCTACTTGGAAAT  
*AtRH24* ACTTGAGTTGTTTCTTCTCTCCGATTTTACATAACCACCGGAAAAACACGAAGCCCT  
*AtRH25* GTACCCTTAAGCCCTAAATTGTGTTTCTTCCACAAAACCTTTTAAACCCCTAAAATAACCGAC  
*AtRH26* TTTTCATGTAACCCCTCTCGACATTTTTCGACTCTTCTTCTCTTATCCTCTAAATCTC  
*AtRH27* TTGAAATGAAATAGAGAGATCTT **TATAAA** CCCTAATGTGACGGTAAACAGTAATACATT  
*AtRH28* ACACATTATACAAAACCTTTTTTGAAGAAAATAACATCTTGGTGTTTTGTATCATCATTT  
*AtRH29* no 5' UTR  
*AtRH30* AATAAAGGAAATTAATTAATTCGCGC **AATAAAAA** GAACAGAAAAATGGTGGGACAAATT  
*AtRH31* no 5' UTR  
*AtRH32* no 5' UTR  
*AtRH33* no 5' UTR  
*AtRH34* TCGTCGTACAAGTAATAAATCAGAACCCTAAAACCCCTTTTCATCTCTCCGGACACTTCT  
*AtRH35* CAGATAATAGGCCCAATAAACAAAATTTACACGCCGTCTAAACCCGATTTCCGACGACCT  
*AtRH36* no 5' UTR  
*AtRH37* AAGGGAGGAGCTTTGAGAATTTGATTGTTGATATCCAAGGATTTGCAAGCTAGTTTTTTTT  
***AtRH38*** AGCTTCAAAGTAAAGCAGCAACTCT **TATTTA** TTTGGGTTTTTCACTTTTTCCCAAAATT  
*AtRH39* ATCTTCTTCTTCTTCTTTCGCTGAACAACAACAACGCTTACAAAACCTTCCGATAAACC  
***AtRH40*** CCCAGCCTCACTTCTTCTTCCAAAAGCCGAACCCAGACCCTGTTCTCTAGTTCCCATCGA  
***AtRH41*** CGGGCCCAATTGTCTTGTTCATCATCTTCGTCTCGCCGCCACTTGCATTAACCAATCC  
***AtRH42*** TTTAGCAAACCTTTTCCACAATTTCCAGACAGTTTTAGGGTTCGGTCTTATTACT  
*AtRH43* no 5' UTR  
*AtRH44* no 5' UTR  
*AtRH45* no 5' UTR  
*AtRH46* ATTTGGAGTGCCCTTTTTGGTGGACCAAATTTTTGGTAAGGGATTGCAAATTCACCCAA  
***AtRH47*** ACAAATTTGAACAATACAAATCAAACCTCGAGTCAGAAAAGATTCAAACAAATGAATCAA  
*AtRH48* no 5' UTR  
*AtRH49* TAGTGGGCTTTTAGGAGAACTGTAGCAAAAACAACAGATTTGCAACAAAATTTTGATGT  
*AtRH50* TTATTTATTTTTTTTTTTTTGGGTCATCTCCTCTCTTATCTCTATGATGACTATACCGTTT  
***AtRH51*** ATAAAAGGTTACGAATTACACAACCTTCAAAGGCTCGTTAAACCCCTAAACCTTGTCTCCGC  
***AtRH52*** TTTGGAAATTAATAAAGTTTCATCGAAATACAAATATCTCTTCTCTTTTCTCTTTTGCTC  
*AtRH53* TTCTTTCTTCTCTCTCTAAATCCCTAATTCTCCAAATTTGGGGTTTTACATTTTCTCT  
*AtRH54, 55* pseudogènes  
***AtRH56*** CCATTCTCAACCCTATCCAGTCTCCCTG **TATATATATA** TTTACGACACCAACCCAGCGTT  
***AtRH57*** CCCAAAATCTGCAAGAGTCTCTCTGGTTCTTCGATTCTTCTCGGTATCTATTCGCCGTCG  
*AtRH58* ACTAAAATTTGACAAAG **TAAAAA** CAGAGTTTTTTATCTGTAAGTAAGAAAAAACAATTC

**Figure 14** Séquence de la région en amont du site apparent de transcription (TSS) des gènes *AtRH*.

Les *AtRH* en gras possèdent un intron en 5' UTR. La séquence des boîtes TATA caractéristiques et modifiées sont surlignées respectivement en noir et gris.

### III. Eléments *cis*-régulateurs et niveaux de transcription

#### A. Classification des *AtRH* en fonction de la structure des régions promotrices

##### 1. Annotation structurale des *AtRH*

Les récents projets de production et séquençage d'ADNc pleine longueur chez *A. thaliana* ont permis d'augmenter et d'améliorer de façon significative les données disponibles pour les transcrits (Haas *et al.*, 2002; Seki *et al.*, 2002c; Seki *et al.*, 2002d; Castelli *et al.*, 2004). Pour caractériser précisément la structure intron-exon et les parties codantes et non codantes des *AtRH*, nous avons aligné leur séquence génomique avec celle de leurs transcrits. Lorsque aucune séquence transcrite n'était disponible, nous nous sommes appuyés sur les prédictions des logiciels NetGene2 et GeneMark HMM et sur la structure de paralogues proches. Au 1er février 2003, les transcrits présents dans les bases de données publiques ont permis de déterminer la région 5' UTR de 46 gènes *AtRH* sur les 56 de la famille. Les *AtRH* ont alors été répartis en trois groupes en fonction de la structure de leur 5' UTR. Le groupe I rassemble les 15 *AtRH* avec un intron dans la région 5' UTR, démontré expérimentalement par la présence de séquences de transcrits dans les bases de données. Leur région 5' UTR est donc composée d'un exon non codant, d'un intron et d'un deuxième exon contenant l'ATG initiateur. La taille des introns varie entre 65 et 649 pb, avec une moyenne de 289 pb. Le groupe II contient les 31 *AtRH* avec une région 5' UTR sans intron. La longueur de la partie exonique de la région 5' UTR des 46 *AtRH*, interrompue ou non par un intron, est similaire :  $123 \pm 98$  pb et  $147 \pm 54$  pb pour les groupes I et II, respectivement. Le groupe III correspond aux 10 *AtRH* dont la région 5' UTR n'est pas connue. Très peu d'EST (0-3) sont disponibles pour ces gènes, indiquant un niveau d'expression transcriptionnel très faible.

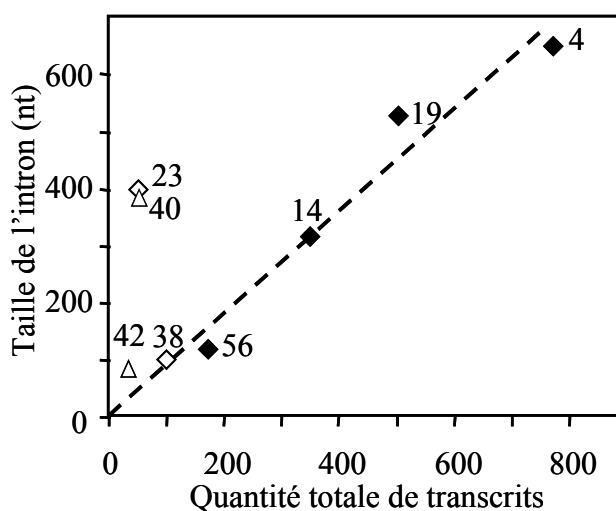
##### 2. Identification de la boîte TATA

Nous avons recherché la boîte TATA dans les 60 pb en amont du site apparent d'initiation de la transcription des gènes *AtRH*, à partir de la séquence consensus TATAWA (Joshi, 1987). Les 10 gènes dont la région 5' UTR n'est pas connue n'ont pas été analysés. En acceptant une substitution d'un A en T ou d'un T en A par motif, une boîte TATA a été identifiée pour 17 gènes sur 46 (Figure 14). Huit gènes présentent une boîte TATA correspondant au consensus et les neuf autres gènes une boîte avec une substitution. La position des boîtes TATA varie de -27 à -50 pb en amont du site apparent d'initiation de la transcription. La majorité des boîtes TATA (15/17), dont les 8 boîtes canoniques identifiées, sont localisée entre -27 et -37 pb.

**Tableau 11 Présence d'une boîte TATA dans la séquence promotrice des gènes *AtRH* selon leur groupe structurel.**

Les chiffres romains désignent les groupes de gènes qui possèdent un intron en 5' UTR (I), qui n'en possèdent pas (II) ou dont la région 5' UTR n'est pas connue (III). Ces groupes sont subdivisés en fonction de la présence d'une boîte TATA consensus ( $I^+$ ,  $II^+$ ), modifiée ( $I^\Delta$ ,  $II^\Delta$ ) ou de l'absence de boîte TATA (I, II). La présence d'une boîte TATA n'a pas été recherchée pour les gènes du groupe III.

Nombre de gènes	TATA consensuelle ( $I^+$ )	TATA modifiée ( $I^\Delta$ )	Sans boîte TATA	Total
Intron (I)	6	2	7	15
Sans intron (II)	2	7	22	31
Sans 5' UTR (III)	-	-	-	10
Total	8	9	29	56

**Figure 15 Influence de la taille de l'intron en 5' UTR sur la transcription des *AtRH* mesurée par PCR quantitative en temps réel.**

Les huit gènes sélectionnés correspondent à l'ensemble des gènes possédant un intron en 5' UTR et dont l'expression transcriptionnelle a été mesurée par PCR quantitative en temps réel. Les gènes pour lesquels une boîte TATA a été identifiée ( $I^+$ ) sont représentés par des losanges noirs et les gènes sans boîte TATA par des triangles blancs (I). Les gènes avec une boîte TATA modifiée sont représentés par des losanges blancs ( $I^\Delta$ ). La quantité totale de transcrits est exprimée en  $10^2$  eg.ng<sup>-1</sup> ARN totaux (Tableau 9). La taille de l'intron est exprimée en nucléotides.

Les groupes I et II définis précédemment ont été subdivisés en fonction de la présence d'une boîte TATA consensus ( $I^+$ ,  $II^+$ ), modifiée ( $I^\Delta$ ,  $II^\Delta$ ) ou de l'absence de boîte TATA (I, II). Il y a un biais de répartition des boîtes TATA (Tableau 11). Les boîtes respectant le consensus sont plus fréquentes lorsqu'un intron est présent dans la région 5' UTR (groupe I). Comme il a été montré que la séquence de la boîte TATA influence l'efficacité de l'expression transcriptionnelle (Starr *et al.*, 1995; Yean & Gralla, 1997), ce biais pourrait indiquer un lien transcriptionnel entre la présence d'une boîte TATA fonctionnelle et celle d'un intron en 5'UTR.

#### B. Groupes structuraux et niveaux d'expression des gènes *AtRH*

Les plus forts niveaux d'expression transcriptionnelle sont observés pour les *AtRH* du groupe  $I^+$  (gènes possédant à la fois un intron dans la région 5' UTR et une boîte TATA), soit *AtRH4*, *14*, *19* et *56*. Plus précisément, nous observons une corrélation directe pour ces *AtRH* entre la taille de l'intron en 5' UTR et la quantité totale de transcrits mesurée par RT-PCR quantitative en temps réel (Figure 15). En effet, la pente de la droite de régression est proche de 1 (0,86;  $r^2 = 0,96$ ). Cette corrélation est aussi observée ( $r^2 = 0,84$ ) lorsque le nombre d'EST disponible pour chaque gène est utilisé comme estimateur de l'abondance des transcrits. Le rôle des introns en 5' UTR dans la régulation de la transcription a été expérimentalement démontré par exemple pour les gènes ubiquitine et les facteurs d'élongation *EF1 $\alpha$*  d'*A. thaliana* (Axelos *et al.*, 1989; Curie *et al.*, 1991; Norris *et al.*, 1993). Les mécanismes impliqués restent inconnus. Un effet secondaire de l'épissage de l'intron, la stimulation de la re-initiation de la transcription par l'ADN polymérase Pol II (Manley, 2002; Dieci & Sentenac, 2003) ou la présence d'informations cryptiques contenues dans l'intron pourraient être envisagés. A notre connaissance, c'est la première fois qu'une corrélation entre la taille de l'intron en 5' UTR et la transcription est observée.

Nos données démontrent que la présence d'un intron en 5' UTR n'est pas suffisante pour prédire un fort niveau d'expression. En effet, les gènes *AtRH23*, *AtRH40* et dans une moindre mesure *AtRH42*, se démarquent significativement de la corrélation. En effet, 11, 10 et 4 fois moins de transcrits ont été mesurés par rapport à ce que la taille de l'intron permettait de prédire. De façon intéressante, ces trois gènes ont un intron en 5' UTR mais aucune boîte TATA canonique. *AtRH40* et *AtRH42* ne possèdent pas de boîte TATA et *AtRH23* présente une boîte modifiée. Ainsi, la corrélation entre la taille de l'intron et la quantité totale de

**Tableau 12** Eléments en amont du site d'initiation de la traduction et niveau d'expression des *AtRH*.

Les gènes sont classés en six groupes en fonction de la présence ou de l'absence d'un intron en 5' UTR (I/II) et/ou d'une boîte TATA canonique ou modifiée (<sup>+</sup>/<sup>Δ</sup>). La colonne « Expression totale » indique la somme des niveaux d'expression mesurée dans les neuf organes étudiés (en  $\text{eg.ng}^{-1}$  ARN totaux). La prédiction de la localisation cellulaire des produits des gènes est celle indiquée dans FLAGdb<sup>++18</sup> (Samson *et al.*, 2002). Lorsque l'adressage est mitochondrial, le numéro du gène est souligné, lorsqu'il est chloroplastique, le numéro du gène est en italique et lorsqu'il est nucléaire, le numéro du gène est suivi d'un n en exposant.

	<i>AtRH</i>	Expression totale	Autres <i>AtRH</i>	Nombre de gènes
<b>Groupe I<sup>+</sup></b>				<b>6</b>
Intron en 5'UTR	04	76543	11, 15	
Boîte TATA	14	35063		
	19	50148		
	56	17252		
<b>Groupe I<sup>Δ</sup></b>				<b>2</b>
Intron en 5'UTR	23	4652		
Boîte TATA modifiée	38	9730		
<b>Groupe I</b>				<b>7</b>
Intron en 5'UTR	40 <sup>n</sup>	4563	41, 47, 51, 52, 57	
	42 <sup>n</sup>	2776		
<b>Groupe II<sup>+</sup></b>				<b>2</b>
Boîte TATA			07 <sup>n</sup> , 27 <sup>n</sup>	
<b>Groupe II<sup>Δ</sup></b>				<b>7</b>
Boîte TATA modifiée	05 <sup>n</sup>	7526	03, 10 <sup>n</sup>	
	08	8145		
	12	627		
	30	6022		
	58	1713		
<b>Groupe II</b>				<b>22</b>
	01	742	<u>09</u> , 13 <sup>n</sup> , 16, 18 <sup>n</sup> , <u>22</u> , 25, 26, 28 <sup>n</sup> , 35, 37, <u>39</u> , 46, 49,	
	02	18233	50, <u>53</u>	
	06	10965		
	20	5710		
	21 <sup>n</sup>	13846		
	24	4500		
	34	923		

<sup>18</sup> <http://genoplante-info.infobiogen.fr/FLAGdb>



transcrits n'est vraie que lorsque ces *AtRH* contiennent aussi une boîte TATA caractéristique. Par ailleurs, le groupe II contient des *AtRH* montrant des niveaux totaux de transcrits entre  $1/4$  et  $1/110$  de celui du gène *AtRH4* (Tableau 12). De fait, en l'absence d'un intron en 5' UTR, le niveau d'expression transcriptionnel couvre une large gamme de valeurs, ce qui reste à expliquer.

## IV. Etude des promoteurs

### A. Caractérisation des séquences promotrices

Afin d'identifier d'autres éléments pouvant expliquer l'expression des gènes *AtRH*, nous avons analysé la séquence promotrice des 56 gènes *AtRH* par deux approches informatiques : une représentation de l'ADN en 3 dimensions et une recherche de motifs par le logiciel RSAT.

Pour ces deux approches, nous avons défini comme région promotrice des *AtRH*, les mille paires de bases situées en amont du codon d'initiation de la traduction. Une fois la position de l'ATG initiateur définie pour chaque *AtRH*, les séquences promotrices ont été extraites automatiquement des fichiers contenant les séquences génomiques. C'est ce jeu de séquences qui a été utilisé pour l'analyse des promoteurs.

### B. Représentation en 3 dimensions

L'approche utilisée au laboratoire provient d'une collaboration avec une équipe de recherche du LIMSI-CNRS. Cette équipe exploite un modèle pour calculer la trajectoire que prendrait l'ADN nu sous la seule contrainte physique de l'enchaînement des bases. L'algorithme permet de représenter les séquences d'ADN par une courbe dans un espace à trois dimensions et de calculer sa courbure.

Au laboratoire, ce logiciel a été utilisé pour rechercher des courbures particulières au niveau des sites d'épissage des introns. Nous l'avons aussi utilisé avec les promoteurs des 56 *AtRH* mais l'allure des courbes obtenues ne nous a pas permis de mettre en évidence de caractéristiques communes. En effet, il est très difficile pour l'œil humain de comparer des tracés en 3D. De plus, le nombre de séquences était trop faible pour effectuer des calculs statistiques.

### C. Analyse des promoteurs et de la région 5' UTR avec le logiciel RSAT

Le logiciel *Regulatory Sequence Analysis Tools* (van Helden *et al.*, 1998) a été développé pour rechercher par une méthode statistique des motifs sur-représentés dans un jeu de séquences. Il est basé sur le principe que les gènes répondant aux mêmes signaux de régulation doivent présenter des sites conservés, correspondants à la zone de fixation du(des) facteur(s) de transcription les régulant. Ainsi une analyse statistique des séquences des régions promotrices d'un groupe de gènes régulés par le même facteur devrait faire apparaître des motifs sur-représentés dans ce groupe de gène par rapport aux séquences promotrices de

l'ensemble du génome. Ce logiciel a été mis au point chez *S. cerevisiae*. Il est actuellement disponible pour un grand nombre de génomes entièrement séquencés soit plus d'une centaine de génomes procaryotes et cinq génomes eucaryotes dont celui d'*A. thaliana*<sup>19</sup>.

Comme les *AtRH* présentent un profil d'expression similaire, nous avons posé l'hypothèse que les gènes de cette famille seraient régulés par le(s) même(s) facteur(s) de transcription. Ainsi, ils forment un ensemble de gènes cohérents avec les critères d'analyse du logiciel RSAT.

### 1. Deux motifs sur-représentés dans le promoteur des *AtRH*

Le promoteur et la région 5' UTR des 56 gènes de la famille *AtRH* ont donc été analysés à l'aide du logiciel RSAT (van Helden *et al.*, 1998). Cette analyse révèle la présence de 2 motifs sur-représentés, AAACCCTA / TAGGGTTT (séquences directe et inverse-complémentaire) et GGCCCA / TGGGCC, avec un index de significativité de 6,02 et 1,98 respectivement.

Le premier motif, AAACCCTA, correspond à la « *telo-box* » ou boîte télo, identifiée dans le promoteur des quatre membres de la famille *AtEF1a* (Axelos *et al.*, 1989) et conservée dans tous les promoteurs des gènes *EF1a* connus chez les Végétaux. Cette boîte a été également décrite dans la région 5' d'autres gènes codant des composants de l'appareil de traduction d'*A. thaliana* (Manevski *et al.*, 1999). La boîte télo est impliquée dans l'activation de l'expression des gènes dans les primordia racinaires, en coopération avec la boîte *tef* ou « *tef-box* » (Manevski *et al.*, 1999; Manevski *et al.*, 2000).

Le second motif, GGCCCA, a d'abord été impliqué dans le contrôle de l'expression des gènes *PCNA* (*Proliferating Cellular Nuclear Antigen*) chez le Riz (Kosugi *et al.*, 1995; Kosugi & Ohashi, 1997; Kosugi & Ohashi, 2002) et *A. thaliana* (Tremousaygue *et al.*, 2003). Il a été décrit dans le promoteur des gènes codant les protéines ribosomiques (Tremousaygue *et al.*, 2003).

### 2. Analyse à l'échelle du génome

#### a. Abondance des boîtes télo et GGCCCA

Nous avons examiné la séquence promotrice des *AtRH* correspondant aux 500 pb en amont du codon d'initiation de la traduction (Tableau 13). Au moins une boîte télo a été identifiée dans

---

<sup>19</sup> <http://rsat.ulb.ac.be/rsat/>

**Tableau 13** Nombre de gènes d'*A. thaliana* présentant au moins une boîte télo, au moins une boîte GGCCA ou au moins un exemplaire de chaque boîte dans les 500 pb en amont du début de la traduction.

Les gènes ont été recherchés dans la base de données FLAGdb<sup>++20</sup> par des mots clés ou par une liste de numéros d'accèsion (italiques). <sup>a</sup> Gènes codant les protéines du cycle cellulaire (Vandepoele *et al.*, 2002).

Fonctions	Nombre total de gènes	Nombre de gènes avec au moins une boîte télo	%	Nombre de gènes avec au moins une boîte GGCCA	%	Nombre de gènes avec les deux boîtes	%
<b>Tous les gènes</b>	<b>24659</b>	<b>3831</b>	<b>16</b>	<b>4036</b>	<b>16</b>	<b>937</b>	<b>4</b>
Protéines ribosomiques	309	197	64	168	54	128	41
<i>AtRH</i>	56	27	48	21	38	12	21
Traduction	62	26	42	19	31	11	18
Elongation	30	11	37	7	23	4	13
Protéasome	52	15	29	27	52	10	19
ARN-t synthétases	45	12	27	9	20	5	11
MADS	96	24	25	6	6	2	2
PPR	470	89	19	133	28	24	5
Protéines du cycle cellulaire <sup>a</sup>	61	9	15	5	8	0	0
Phosphatases	205	28	14	28	14	2	1
Transcription	472	60	13	32	7	7	1
Histones	58	7	12	21	41	3	6
Facteurs MYB	184	20	11	7	4	2	1
Sérine/thréonine kinases	131	11	8	17	13	4	3
Métabolisme	2649	220	8	329	12	47	2
Cytochrome P450	245	12	5	18	7	0	0
<b>Toutes les catégories</b>	<b>5125</b>	<b>768</b>	<b>15</b>	<b>847</b>	<b>17</b>	<b>261</b>	<b>5</b>

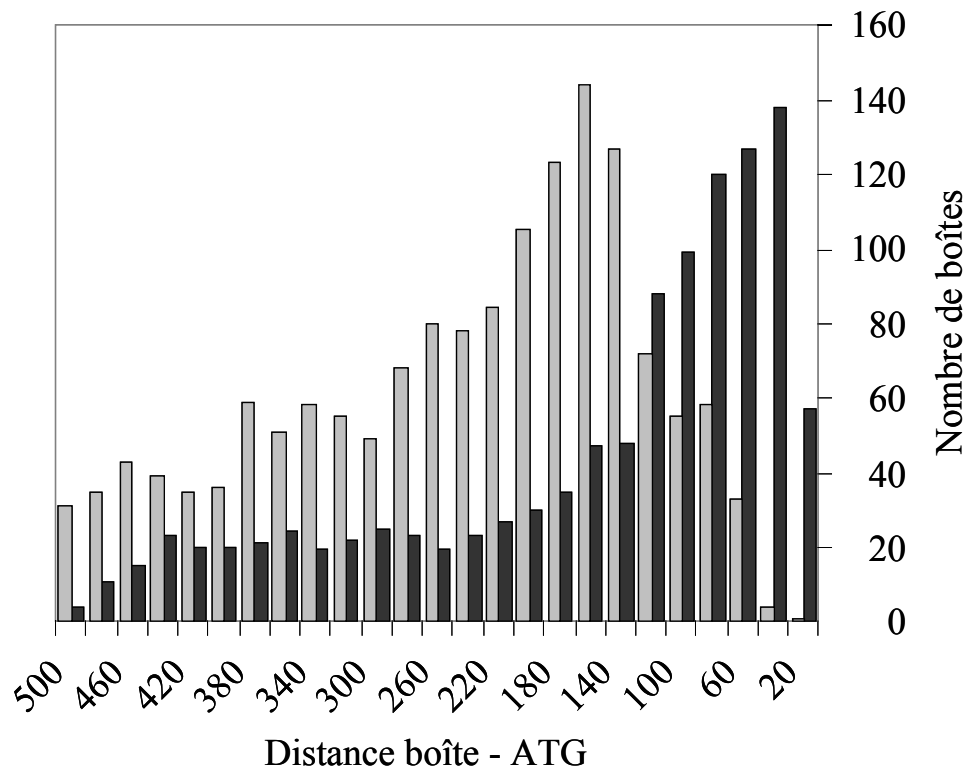
<sup>20</sup> <http://genoplante-info.infobiogen.fr/FLAGdb>

27 *AtRH* (48%) et au moins une boîte GGCCCA dans 21 *AtRH* (38%). Douze *AtRH* (21%) présentent au moins un exemplaire de chacune des deux boîtes. La même étude a été réalisée sur l'ensemble des 24,659 gènes d'*A. thaliana* (prédictions TIGR/GenBank) à l'exception des gènes des régions répétées comme les transposons. Environ 16% (3831) des gènes contiennent une boîte télé, environ 16% (4036) présentent une boîte GGCCCA et 4% (937) possèdent les deux boîtes télé et GGCCCA (Tableau 13) dans les 500 pb en amont de leur codon d'initiation de la traduction.

La différence de fréquence de la présence des deux boîtes entre l'ensemble des gènes d'*A. thaliana* et au sein des *AtRH* suggère que ces deux boîtes pourraient être présentes de façon préférentielle dans certaines familles de gènes. Ainsi, une recherche exhaustive dans les 500 pb en amont du codon d'initiation de la traduction de plusieurs catégories fonctionnelles de gènes ou de grandes familles de paralogues d'*A. thaliana* a été réalisée (Tableau 13).

Les gènes correspondant à chaque catégorie ont été recherchés à l'aide de mots clé dans la base de donnée FLAGdb<sup>++</sup>, sauf les *AtRH*, déjà identifiées et dont la liste des numéros d'accèsion était disponible au laboratoire, et les protéines du cycle cellulaire qui correspondent à l'ensemble défini par Vandepoele et ses collaborateurs (Vandepoele *et al.*, 2002). Parmi ces gènes, représentant environ 25% du génome, 15% contiennent une boîte télé, 17% une boîte GGCCCA et 5% les deux. La proportion de gènes présentant l'une ou l'autre des boîtes varie beaucoup selon le groupe de gènes considéré. La boîte télé est présente dans 5% des gènes de la famille des *P450* et dans 64% des gènes de la famille des protéines ribosomiques. La boîte GGCCCA est présente dans 4% des gènes de la famille des facteurs *MYB* et dans 54% des gènes de la famille des protéines ribosomiques. De plus, l'abondance relative de chaque boîte est différente selon les catégories fonctionnelles. Par exemple, les familles des histones et des phosphatases présentent une fréquence de boîte télé similaire : 12% et 14 % des gènes respectivement, alors que l'occurrence de la boîte GGCCCA est respectivement de 41% et 14% des gènes pour ces deux familles. Inversement, les familles des *AtRH* et des histones présentent une fréquence de boîte GGCCCA similaire : 41% et 38% des gènes, respectivement et des fréquences de boîte télé différentes : 48% et 12% des gènes.

Les gènes classés dans les catégories fonctionnelles telles que traduction, élongation et protéasome présentent une proportion élevée de gènes possédant le motif télé : 30 à 40% environ. Par contre, les gènes classés dans les catégories transcription et cycle cellulaire se rapprochent beaucoup plus de la moyenne avec 13% et 15% de gènes portant un motif télé. Et enfin, seulement 8% des gènes impliqués dans le métabolisme présentent un motif télé.



**Figure 16 Répartition des boîtes télomériques et GGCCA dans la région promotrice des 937 gènes d'*A. thaliana* possédant les deux boîtes.**

Les barres indiquent le nombre de boîtes télomériques (en noir) et de boîtes GGCCA (en gris) présentes dans les 500 pb en amont du codon d'initiation de la traduction, par intervalles de 20 nucléotides.

### b. Localisation et répartition des boîtes dans les promoteurs

Après avoir analysé l'abondance des boîtes télo et GGCCCA, nous avons étudié leur position dans l'ensemble des séquences promotrices du génome d'*A. thaliana*. La figure ci-contre illustre la répartition des deux boîtes dans les 937 gènes contenant au moins chacune des deux boîtes dans les 500 pb en amont du codon d'initiation de la traduction. Les boîtes télo et GGCCCA n'occupent pas la même position (Figure 16).

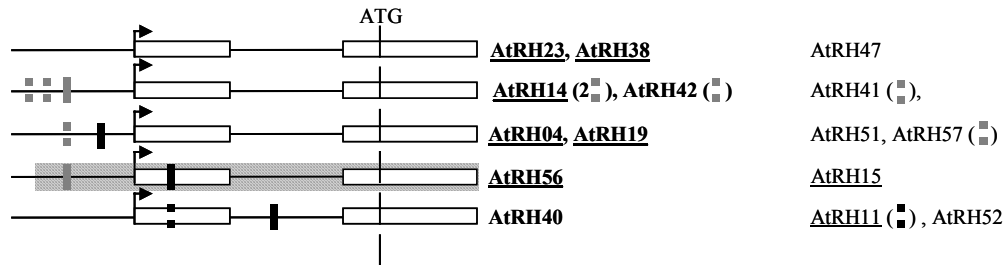
La boîte télo est positionnée préférentiellement en aval de la boîte GGCCCA, dans les 150 pb en amont du codon d'initiation de la traduction. Dans cette région, il y a 4 à 5 fois plus de boîtes télo que dans les autres régions intergéniques et génomiques. En effet, au-delà de 500 pb, la fréquence d'apparition du motif est la même quelle que soit la distance. Autrement dit, la présence du motif télo au-delà de ces 500 nucléotides en amont de l'ATG serait aléatoire et non soumise à une pression de sélection. Cette configuration de la répartition du motif télo laisse supposer un rôle biologique aux motifs télo proches de l'ATG.

La répartition des boîtes GGCCCA est différente car les boîtes GGCCCA sont rarement présentes dans les 120 premières paires de bases en amont du codon d'initiation de la traduction mais sont principalement détectées dans une région située entre -120 et -270 pb. La boîte GGCCCA est souvent présente sous forme de diades ou de triades. La distance entre deux boîtes GGCCCA n'est pas distribuée au hasard car 9 paires de boîtes sur 12 sont séparées par 10 à 12 pb ou 21 à 25 pb.

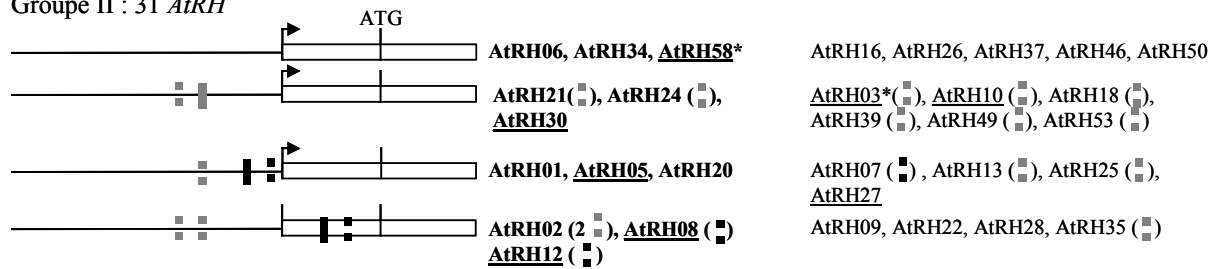
Trois *AtRH* contenant un long intron dans la région 5' UTR possèdent une boîte télo relativement proche du début de leur transcription, mais au-delà de la région de 500 pb en amont du codon d'initiation de la traduction. Ces données suggèrent une conservation de la position des boîtes télo et GGCCCA en amont du site d'initiation de la transcription plutôt que de celui de la traduction. Ainsi, les boîtes télo des gènes dont on connaît la région 5' UTR sont localisées à  $-37 \pm 20$  pb du site apparent de leur transcription alors que les boîtes GGCCCA le sont à  $-112 \pm 47$  pb. Cependant, quelques boîtes télo ont été identifiées dans les parties exoniques et/ou introniques de la région 5' UTR de certaines *AtRH* (Figure 17).

Le groupe III des *AtRH* sans transcrits associés à la région 5' UTR correspond aux gènes pour lesquels peu d'EST sont disponibles. Les gènes de ce groupe contiennent fréquemment les boîtes télo et GGCCCA (8/10) et même des diades des deux boîtes, mais nous n'avons jamais identifié de boîte TATA. Dans ce groupe, les boîtes télo sont positionnées à  $-82 \pm 37$  bp et les boîtes GGCCCA à  $-174 \pm 38$  pb en amont du début de la traduction. Ces distances suggèrent que ces gènes ne contiennent probablement pas d'intron dans leur région 5' UTR.

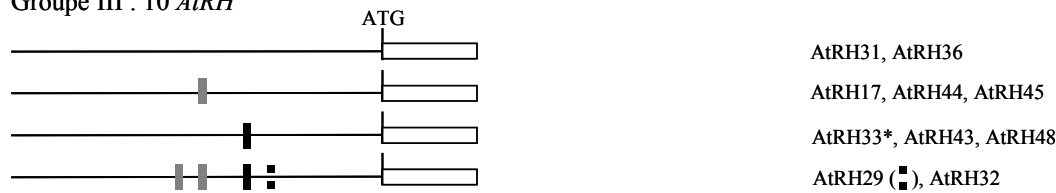
Groupe I : 15 *AtRH*



Groupe II : 31 *AtRH*



Groupe III : 10 *AtRH*



**Figure 17** Disposition de certains éléments *cis*-régulateurs de la famille des *AtRH* dans la partie 5' en amont du codon d'initiation de la traduction.

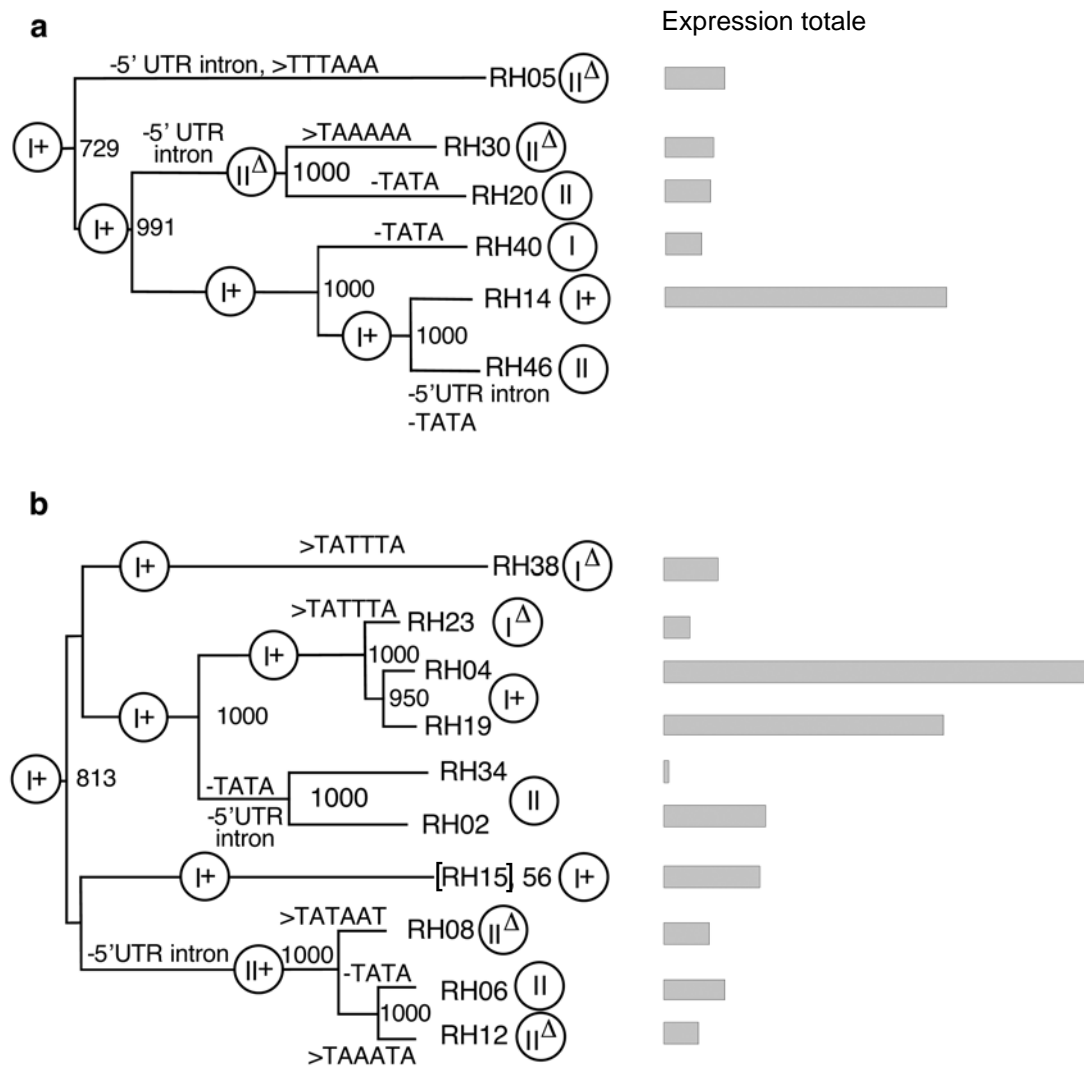
Les exons sont symbolisés par des rectangles blancs. Les introns et les régions non transcrites par des lignes horizontales. Le début de l'ADNc le plus long en 5' est indiqué par une flèche et l'ATG initiateur par une ligne verticale. Les boîtes télo sont représentées en noir et les boîtes GGCCCA en gris. Lorsqu'une boîte est présente dans l'ensemble des gènes du groupe, elle est signalée par un trait plein. Les traits pointillés indiquent la présence d'une boîte pour au moins l'un des gènes du groupe. Dans ce cas, le symbole correspondant figure entre parenthèse après le nom du gène concerné. Si plusieurs boîtes existent, leur nombre est précisé dans la parenthèse. Les hachures délimitent une région des gènes d'un sous-groupe identiques à plus de 95%. Un nom de gène souligné signale la présence d'une boîte TATA située à environ 30 nucléotides avant le début de la transcription. Le nom des *AtRH* dont la transcription a été mesurée par RT-PCR quantitative en temps réel sont écrits en gras. L'astérisque indique une prédiction d'adressage chloroplastique de la protéine.



Nos données s'ajoutent à une récente étude réalisée par Tremousaygue *et al.* (2003) montrant une association topologique conservée entre les boîtes télomériques et (A/G)GCCCA dans les gènes codant les protéines ribosomiques.

### 3. Lien entre expression et boîtes

Même si les boîtes télomériques et GCCCA présentent une fréquence et une répartition particulières dans la région promotrice des *AtRH*, il n'a pas été mis en évidence de lien entre le niveau d'expression et la présence ou l'absence de ces boîtes. Les profils d'expression ayant été établis à partir d'organes entiers, il est possible qu'une expression différentielle ait été noyée dans l'ensemble des cellules de l'organe. Par ailleurs, les études fonctionnelles de la boîte télomérique des gènes *EFL1* d'*A. thaliana* ont montré la nécessité d'enlever l'intron « leader » situé dans la région 5' UTR pour déterminer l'influence de la boîte télomérique sur l'activité du promoteur (Axelos *et al.*, 1989). Il est donc possible que le rôle des boîtes télomériques et GCCCA sur l'expression des *AtRH* soit masqué par les composants majeurs du contrôle de l'expression à savoir, la présence d'un intron en 5' UTR et d'une boîte TATA.



**Figure 18** Arbre de distance des séquences protéiques de deux groupes d'*AtRH* superposée à la reconstitution de l'évolution des éléments en amont du site d'initiation de la traduction et niveau d'expression totale des gènes *AtRH*.

-TATA signale la perte de la boîte TATA. -5' UTR intron indique la perte de l'intron en 5' UTR. >TATTTA montre la modification de la séquence de la boîte TATA. La valeur des *bootstraps* est calculée à partir de 1000 répétitions. Les protéines codées par les gènes *AtRH15* et *AtRH46* sont identiques. Les groupes sont indiqués dans les cercles. I : intron en 5' UTR ; II : absence d'intron en 5' UTR ; + : boîte TATA ;  $\Delta$  : boîte TATA modifiée. Les barres grises représentent l'expression totale mesurée par RT-PCR quantitative en temps réel. La longueur des barres est proportionnelle à l'expression. L'expression du gène *AtRH15* n'est pas indiquée car elle n'a pas été mesurée.

## V. Groupes d'orthologues et transcription

Suite à la vaste prolifération de la famille chez *A. thaliana* (Anantharaman *et al.*, 2002), les groupes d'orthologues entre les *AtRH* et les ARN hélicases à boîte DEAD de *S. cerevisiae* (*ScRH*) contiennent jusqu'à cinq *AtRH* associées avec une ou deux *ScRH*. La figure 17 met en relation la présence d'éléments régulateurs dans les régions promotrices et le niveau d'expression des gènes de plusieurs de ces groupes.

Le groupe d'orthologues caractérisé par *ScRH-DBP2* contient les 5 gènes *AtRH14*, *20*, *30*, *40* et *46* (Figure 18a). Les arbres de distance établis par Boudet *et al.* (2001) indiquent qu'un premier élément de duplication a généré les ancêtres de deux sous-groupes de gènes, dupliqués plus récemment : *AtRH20* et *30* d'un côté et *AtRH14*, *40* et *46* de l'autre. Cette reconstruction phylogénétique est renforcée par le fait que les deux sous-groupes présentent une structure de gène complètement différente. En effet, *AtRH20* et *30*, mais pas les gènes *AtRH14*, *40* et *46* possèdent un intron dans la région codante à une position conservée dans tous les organismes eucaryotes dont le génome est complètement séquencé ou presque (Boudet *et al.*, 2001). Ces cinq gènes présentent quatre structures promotrices différentes : *AtRH14* possède un intron en 5' UTR et une boîte TATA caractéristique ; *AtRH40* possède également un intron 5' UTR mais pas de boîte TATA ; *AtRH46* et *AtRH20* ne possèdent pas d'intron en 5' UTR ni de boîte TATA ; *AtRH30* ne possède pas d'intron en 5' UTR mais présente une boîte TATA modifiée. *AtRH20* et *30* présentent un niveau d'expression similaire, environ 6 fois moins qu'*AtRH14*. Les gènes *AtRH40* et *46* présentent une expression encore plus faible. Ainsi, dans un groupe de gènes récemment dupliqués, nous observons une modification rapide de la région promotrice, corrélée à un changement concomitant de l'activité transcriptionnelle. La reconstruction des événements qui ont conduit à la situation actuelle des gènes *AtRH14*, *20*, *30*, *40* et *46* suggère qu'après chaque événement de duplication d'un gène ancestral possédant un intron en 5' UTR et une boîte TATA caractéristique, l'un des paralogues généré diverge de l'autre en perdant soit la boîte TATA, soit l'intron en 5' UTR. La seule exception est observée pour la duplication la plus récente du groupe, ayant généré les gènes *AtRH14* et *AtRH46*. Ce dernier événement implique la perte des deux éléments régulateurs : l'intron en 5' UTR et la boîte TATA, conduisant à un gène fortement exprimé et à un paralogue très faiblement exprimé.

Une reconstruction similaire des événements évolutifs peut être proposée pour d'autres branches de l'arbre de distance des *AtRH* (Figure 18b). La plupart du temps, un seul événement se produit après chaque duplication et, à chaque fois, il s'agit de la perte d'un élément régulateur. Dans les duplications plus anciennes, par exemple celle à l'origine des

membres de la branche terminale contenant *AtRH02* et *34*, la boîte TATA et l'intron en 5'UTR ont été perdus, suggérant la perte d'un gène dupliqué possédant l'un de ces deux éléments régulateurs.

## Discussion

### I. Quantification de la teneur des ARNm par RT-PCR quantitative en temps réel

#### A. Recommandations

La PCR quantitative en temps réel permet de mesurer la teneur d'une molécule d'acide nucléique particulière dans un extrait d'ADN. La sensibilité de cette technique rend nécessaire de prendre un certain nombre de précautions, notamment lors de l'étude des transcrits.

Tout d'abord, comme les transcrits dosés par PCR quantitative sont souvent peu abondants, l'absence de contamination par de l'ADN génomique dans l'échantillon d'ARN est cruciale. Il est donc important d'obtenir un échantillon d'ARN exempt d'ADN par le biais, par exemple, de la dégradation de l'ADN résiduel par une ADNase « *RNase free* » et de le vérifier en testant par exemple l'absence d'amplification par PCR d'un intron d'un gène ubiquitaire.

Par ailleurs, l'importance des familles de gènes dans le génome d'*A. thaliana* et celui des autres plantes influence la stratégie de sélection des amorces pour la PCR quantitative en temps réel. En effet, certains paralogues présentent des séquences si proches qu'un couple d'amorces pourrait amplifier deux gènes. Ainsi, lors de la sélection des amorces, l'attention doit être portée sur la spécificité des couples vis-à-vis de leur gène cible. Pour cela, le recensement de l'ensemble des gènes formant la famille étudiée est primordial.

D'un point de vue technique, la présence d'une gamme-étalon complète sur chaque plaque permet d'une part, de contrôler la reproductibilité des mesures et d'autre part, de s'affranchir de la variation de l'efficacité d'un couple d'amorces à l'autre, autorisant ainsi la comparaison du niveau d'expression de deux gènes. Dans le même but, les couples d'amorces doivent présenter une efficacité d'amplification homogène au plus proche de 100%.

Il a été montré l'intérêt de réaliser la mesure du même gène de référence sur chaque plaque, en parallèle du(des) autre(s) gène(s) mesuré(s), comme témoin positif de la réaction de PCR quantitative et contrôle de la qualité de la gamme et des échantillons.

Enfin, il est judicieux de tester plusieurs quantités différentes de chaque échantillon pour s'assurer de la linéarité de la gamme des mesures. Les mesures retenues doivent, pour être comparées, être proportionnelles à la quantité initiale d'échantillon.

### B. La normalisation des résultats

La normalisation des niveaux de transcrits permet de faciliter la comparaison des résultats d'un échantillon à l'autre. D'après Bustin (2002), la normalisation la plus appropriée est de ramener au nombre de cellules ou à la quantité totale d'ARN. J'ai retenu cette deuxième proposition pour présenter mes résultats.

Lorsque l'étape de RT est réalisée avec des hexamères aléatoires, de nombreux auteurs utilisent comme référence normalisatrice le niveau des ARNr 18S (Itoh *et al.*, 2001; Goda *et al.*, 2002; Klok *et al.*, 2002; Teramoto *et al.*, 2002; Hammond *et al.*, 2003; Liu *et al.*, 2003; Ogawa *et al.*, 2003; Shimada *et al.*, 2003; Stasolla *et al.*, 2003) et 28S (Choi *et al.*, 2002). Cependant, les ribosomes sont très abondants dans la cellule, et la normalisation du niveau de transcrits par rapport aux ARNr, surtout quand il s'agit de gènes faiblement exprimés, risque d'entraîner un biais. De plus, comme les ARNr et les ARNm sont transcrits par différentes ARN polymérase, leur niveau sont indépendants et peuvent varier dans certains mutants ou conditions physiologiques (Bustin, 2002).

Les gènes de maintenance ne sont pas plus appropriés, sauf cas particuliers, car leur expression varie trop souvent d'une condition à l'autre. Par exemple, il a été montré que le niveau des transcrits du gène d'actine *ATC2* est réduit de 50% quand les plants d'*A. thaliana* sont soumis à un choc thermique en les passant de la température ambiante à 37°C alors qu'aucune variation n'est observée en les transférant à 44°C (Volkov *et al.*, 2003). Par contre, le niveau de transcrits du gène codant la protéine ribosomique L23 est réduit lors du passage à 44°C (Volkov *et al.*, 2003). Les gènes de maintenance n'adoptent donc pas le même comportement face au même stress. Par conséquent, il est nécessaire de vérifier si l'expression du gène constitutif choisi n'est pas modifiée par les traitements réalisés, comme l'ont fait par exemple Harari-Steinberg *et al.* (2001) pour le gène d'ubiquitine UBQ10 pendant la croissance de jeunes plantes d'*A. thaliana* à la lumière et à l'obscurité. Pourtant, de nombreux auteurs normalisent par rapport à des gènes « constitutifs » par exemple des gènes codant des protéines du cytosquelette comme l'actine (Charrier *et al.*, 2002; Zheng *et al.*, 2002 et d'autres; Chae *et al.*, 2003; McGinnis *et al.*, 2003) et la tubuline  $\beta$  (Mladek *et al.*, 2003) ou encore le facteur d'initiation de la traduction eIF4A1 (Bonaventure *et al.*, 2003), le facteur d'élongation EF1 $\alpha$  (Collakova & DellaPenna, 2003; Foucher *et al.*, 2003) et les gènes

d'ubiquitine *UBQ2* (Itoh *et al.*, 2001) ou *UBQ10* (Harari-Steinberg *et al.*, 2001; Wang *et al.*, 2003b). Dans une étude de la famille des transporteurs de nitrate *NTR*, Orsel *et al.* (2002a) ont utilisé le gène *AtAPT1* (adénine phosphoribosyl transférase 1) pour normaliser leurs résultats. Vandesompele *et al.* (2002) proposent dans une discussion sur la normalisation des résultats de PCR quantitative d'utiliser plusieurs gènes. Jacobs *et al.* (2003) ont illustré l'utilisation simultanée de trois gènes constitutifs pour normaliser l'expression de gènes de glucane synthases dans des plants d'*A. thaliana* transformées par une construction RNAi.

La majorité des ARN hélicases présentent un profil similaire et peu variable. Ces gènes sont donc de bons candidats de contrôle interne. De plus leurs niveaux d'expression varient de moyen à très faible par rapport à d'autres gènes de maintenance comme *EF1 $\alpha$* , permettant de choisir un gène *AtRH* exprimé à un niveau proche du gène à normaliser, évitant l'écueil des ARNr, trop abondants.

### C. Sensibilité et spécificité

La diversité des systèmes de détection, des différents modèles de machines, des différentes méthodes de calculs et de normalisation des résultats, rend difficile la comparaison des résultats obtenus par différentes équipes. De plus, la plupart des auteurs donnent des valeurs relatives par rapport à un gène contrôle ou à une condition de référence. Or, seule l'indication des valeurs absolues permet de comparer les mesures des différentes expériences. Bustin (2002) insiste sur le fait qu'il reste peu d'améliorations à apporter à la technique de PCR quantitative mais que le manque d'homogénéité dans le mode opératoire augmente la difficulté de comparer les résultats entre différentes équipes. Il pose la question de savoir combien la RT-PCR quantitative est réellement quantitative, reproductible et informative, question que nous avons voulu aborder dans notre travail.

#### 1. Sensibilité

La gamme des mesures possibles en PCR quantitative s'étend de quelques dizaines de molécules à plusieurs millions, couvrant une amplitude de  $10^5$  à  $10^7$  selon les auteurs. L'étendue de cette gamme couvre la majeure partie des conditions rencontrées en physiologie végétale. Par exemple, il est possible de mesurer des variations de gènes inductibles comme le gène *PR1* (Schenk *et al.*, 2003), fortement induit lors d'attaques par des pathogènes et utilisé comme marqueur en phytopathologie.

Au cours de mon travail, j'ai pu évaluer la spécificité et la sensibilité de la PCR quantitative en temps réel. Dans les conditions dans lesquelles je me suis placée, la sensibilité de la PCR quantitative est de l'ordre de  $20 \text{ eg.ng}^{-1}$  ARN totaux. Cet ordre de grandeur de la sensibilité est cohérent avec ceux présentés dans la littérature. Dans leur étude sur la famille des gènes *AtARI*, Mladek *et al.* (2003) donnent comme limite de sensibilité  $5 \text{ copies.ng}^{-1}$  soit environ  $10 \text{ eg.ng}^{-1}$  ARN totaux. Pour cette famille, la limite de la sensibilité de la technique est atteinte : l'expression de seulement la moitié des gènes est mesurable par PCR quantitative en temps réel. Pour l'autre moitié, le signal enregistré se situe au niveau du bruit de fond et ne permet pas de trancher entre absence d'expression ou faible expression. La sensibilité de la PCR quantitative en temps réel permet l'étude de gènes très faiblement exprimés.

Il faut rappeler que les mesures à la limite de la sensibilité, comme les autres, ne sont valables que lorsque la quantité de transcrits mesurée est proportionnelle à la quantité d'échantillon mesurée. De plus, bien que la sensibilité soit, à première vue, uniquement limitée par le dépôt d'une quantité suffisante de matrice pour réaliser la mesure, un excès de matrice peut être tout aussi préjudiciable. En effet, comme la PCR est une réaction enzymatique, elle est soumise à l'inhibition par excès de substrat. En déposant une quantité trop importante de matrice, l'amplification risque de ne pas se dérouler dans des conditions optimales. Par ailleurs, des inhibiteurs de PCR peuvent être présents dans les échantillons dans une proportion variable. Dans ce cas également, plus la quantité d'échantillon déposée est importante, plus l'amplification de la PCR risque d'être inhibée. Nous avons observé ce phénomène en comparant différents extraits d'ADN génomique.

Théoriquement, il n'y a pas de limite maximale de la mesure de l'expression, car la dilution de l'échantillon jusqu'à des valeurs mesurables est toujours possible.

## 2. Spécificité

La spécificité d'une technique se définit par sa capacité à générer le moins possible de résultats « faux positifs ». Dans notre cas, cela correspond à obtenir une valeur supérieure au bruit de fond de la mesure d'une expression nulle. Un vrai résultat négatif correspond à une valeur appartenant au bruit de fond lorsqu'un gène n'est pas exprimé. Parmi mes résultats, seul le gène *SUPI* présente dans certains échantillons des mesures se confondant avec le bruit de fond. Pour estimer la spécificité de la PCR quantitative en temps réel, j'ai comparé mes résultats avec ceux de la littérature. Il a été montré par hybridation *in situ* et par détection de l'expression de gènes rapporteurs que l'expression du gène *SUPI* est limitée à certains tissus de la fleur et de la graine (Sakai *et al.*, 1995; Meister *et al.*, 2002; Ito *et al.*, 2003). Par RT-



PCR quantitative, son expression a été mesurée uniquement dans les bourgeons floraux, les fleurs et les siliques. Dans les autres organes, les mesures obtenues se confondent au bruit de fond (Charrier *et al.*, 2002; Mingam *et al.*, 2004). Cet exemple montre que la PCR quantitative est au moins aussi spécifique que les autres méthodes d'études de l'expression transcriptionnelle.

**Tableau 14 Valeurs minimale et maximale des mesures obtenues par RT-PCR quantitative en temps réel au cours de ce travail.**

Les valeurs sont exprimées en équivalents-génome par nanogramme d'ARN totaux.

Expression minimale	20 eg.ng <sup>-1</sup>
Expression maximale	85 000 eg.ng <sup>-1</sup>
Bruit de fond	< 20 eg.ng <sup>-1</sup>

#### D. Comparaison de la RT-PCR quantitative au *Northern blot*

##### 1. Sensibilité

Lorsque les gènes sont exprimés à de très faibles niveaux, la méthode de RT-PCR est souvent préférée au *Northern blot*. Certains gènes *LeACS*, au départ considérés comme non exprimés au vu des résultats d'expériences de *Northern blot* ou de protection à la ribonucléase, ont montré un niveau extrêmement bas de transcription, et pourtant mesurable par RT-PCR en temps réel (Balbi & Lomax, 2003).

Lorsque les gènes à étudier sont suffisamment exprimés et leur séquence suffisamment différente, il peut être plus avantageux de réaliser l'analyse de leur expression par *Northern blot* plutôt que par RT-PCR quantitative, du moins si l'on se place dans une perspective marchande, tenant compte du coût de l'expérience. Kursteiner *et al.* (2003) ont fait ce choix pour étudier l'expression des gènes *AtALDH* en conditions d'anoxie chez *A. thaliana*. Par contre, ils ont utilisé la RT-PCR quantitative en temps réel pour étudier l'expression des gènes *PDC* dont l'expression de trois d'entre eux n'est pas détectable par *Northern blot*.

Comme l'analyse de l'expression par PCR quantitative demande de faibles quantités d'ADNc (10 à 200 ng comparés à environ 10 à 20 µg pour un *Northern blot*), il est possible de suivre par RT-PCR quantitative l'expression de nombreux gènes à partir d'une simple extraction d'ARN, même si la quantité de matériel est limitée. Lake et Willows (2003) ont montré que la

PCR quantitative en temps réel nécessite 10 à 100 fois moins d'ARN qu'un *Northern blot*, pour obtenir des résultats d'expression similaires.

## 2. Spécificité

Lorsque les gènes présentent des séquences très proches, le phénomène d'hybridation croisée peut rendre le *Northern blot* totalement inapproprié pour l'étude de l'expression. Montrichard *et al.* (2003) ont ainsi remarqué que le signal plus fort observé en *Northern blot* par rapport à celui observé en RT-PCR quantitative était dû à de l'hybridation croisée entre leur sonde et plusieurs transcrits appartenant à la même famille de gènes.

La RT-PCR quantitative peut venir compléter une analyse par *Northern blot*, notamment lorsque les séquences sont très proches (plus de 70% d'identité). Après une analyse de l'expression par *Northern blot*, Brown *et al.* (2003b) ont utilisé la RT-PCR quantitative pour distinguer l'expression du gène de réponse aux pathogènes *PDF1-2* chez *A. thaliana* de celle de trois autres homologues.

## 3. Limites et complémentarités

Sauf dans de rares cas, la PCR quantitative en temps réel ne permet pas de détecter les événements d'épissage alternatif. En effet, les amorces utilisées pour la PCR quantitative risquent d'amplifier sans distinction les deux transcrits. Comme les fragments amplifiés en PCR quantitative sont courts (50 à 200 pb) et souvent à l'intérieur d'un même exon, l'épissage alternatif ne sera pas détecté et l'expression mesurée correspondra à la somme de la teneur des deux transcrits. Dans le cas où au moins l'une des amorces serait localisée à l'intérieur de la partie épissée, seul l'un des transcrits est amplifié. Lorsque les amorces se trouvent de part et d'autre du site d'épissage, il est possible que les deux produits soient amplifiés. Dans ce cas, la courbe de fusion indiquera la présence de deux produits PCR de  $T_m$  différentes, permettant d'émettre l'hypothèse d'un épissage alternatif. Cependant, il peut aussi arriver que seul le transcrit le plus court soit amplifié, ne permettant pas non plus le diagnostic d'un épissage alternatif. Ainsi, le *Northern blot* peut apporter des informations complémentaires importantes.

Par contre, si un épissage alternatif est mis en évidence par *Northern blot*, il est possible de suivre la fréquence de cet événement par RT-PCR quantitative en choisissant les couples d'amorces appropriés (Halterman *et al.*, 2003).

Par ailleurs, lorsqu'une dégradation partielle des transcrits se produit, comme par exemple lors des études d'extinction de gènes (*gene silencing*), la RT-PCR quantitative est peu indiquée. En effet, basée sur l'amplification de courts fragments, la RT-PCR quantitative peut conduire à la comptabilisation de transcrits en cours de dégradation au moment de la mesure de l'expression. Dans ce cas et celui de l'épissage alternatif, la RT-PCR quantitative en temps réel ne doit pas être le seul outil utilisé pour étudier l'expression, et les données devraient être complétées par des analyses par *Northern blot*.

Il est extrêmement important de connaître l'ensemble des gènes d'une famille pour garantir la spécificité de l'analyse. Chez *A. thaliana* et les autres espèces dont le génome est entièrement séquencé, le recensement exhaustif des membres d'une même famille est assez aisément réalisable. Par contre, pour les génomes partiellement déchiffrés comme ceux de nombreuses espèces d'intérêt agronomique, il est difficile de ne pas avoir omis un gène. Chez le pois (*Pisum sativum*), la famille des NADPH/NADP-thiorédoxines TRX h compte au moins quatre membres. En réalisant en parallèle un *Northern blot* et une analyse de l'expression par RT-PCR quantitative en temps réel, Montrichard *et al.* (2003) ont étudié l'expression des gènes *PsTRX h3* et *PsTRX h4*. Cependant, l'analyse du *Northern blot* réalisé avec la sonde correspondant au gène *PsTRX h3* laisse supposer l'existence de deux transcrits de tailles différentes codés par ce gène (épissage alternatif) ou l'existence d'un cinquième gène dans cette famille. Ainsi, le *Northern blot* apporte des informations complémentaires aux résultats de RT-PCR quantitative en temps réel.

Enfin, il faut garder à l'esprit que les analyses de RT-PCR quantitative sont réalisées en général sur des organes entiers composés de différents tissus. Même si Philippar *et al.*, (2003) ont réussi à comparer les profils d'expression de gènes codant des protéines formant des canaux à potassium au niveau de l'épiderme, du mésophyle et des tissus vasculaires des feuilles de Maïs par RT-PCR quantitative, des études de fusions de promoteurs à des gènes rapporteurs peuvent être nécessaires pour compléter la description du profil d'expression (Berger *et al.*, 2002; Reintanz *et al.*, 2002). La technique de microdissection au laser (*Laser Capture Microdissection*) permet d'isoler certaines cellules à partir de coupes de tissus (Emmert-Buck *et al.*, 1996). Des protocoles spécifiques permettent ensuite d'extraire l'ADN et les ARN de ces cellules. Chez les mammifères, cette technique est employée en association avec la RT-PCR quantitative en temps réel (Bustin, 2002). Elle a été appliquée aux plantes pour réaliser des banques de transcrits spécifiques de certains types cellulaires comme le

phloème chez le Riz (Asano *et al.*, 2002; Kerk *et al.*, 2003). Cette technique a été combinée avec une analyse par *microarrays* pour identifier des gènes spécifiques de cellules épidermiques et vasculaires chez le Maïs (Nakazono *et al.*, 2003).

#### E. Une technique complémentaire aux *microarrays*

Les *microarrays* permettent de regrouper les gènes sur la base de leur expression et d'identifier ainsi des gènes co-régulés ou présentant un même profil d'expression. La PCR quantitative en temps réel est la principale technique utilisée pour valider les résultats de *microarrays*. Rajeevan *et al.* (2001) ont montré que la validation dépend de l'intensité du signal d'hybridation. Plus le signal enregistré en *microarrays* est fort, plus il est probable de confirmer le changement d'expression du gène par PCR quantitative, surtout si le changement est supérieur à 4 fois entre les deux conditions d'hybridation. Ces auteurs notent également que seule la nature du changement (augmentation ou diminution) est validée, mais pas son amplitude. En effet, la PCR quantitative donne en général des facteurs de variations plus grands que ceux observés en *microarrays* (Rajeevan *et al.*, 2001; Klok *et al.*, 2002; Wang *et al.*, 2003b; Wurmbach *et al.*, 2003). D'une part, les phénomènes d'hybridation croisée peuvent se produire au cours des expériences de *microarrays* (Schenk *et al.*, 2003). Ils expliqueraient une partie des différences observées entre les valeurs obtenues par *microarrays*, notamment dans le cas d'hybridation à de longs fragments d'ADN (ADNc, EST), et celles obtenues par RT-PCR quantitative. D'autre part, la saturation du signal est atteinte beaucoup plus rapidement avec les *microarrays* qu'avec la PCR quantitative. Enfin, les *microarrays* ne permettent d'étudier que l'expression des gènes les plus exprimés. Ainsi, la PCR quantitative est plus sensible et fiable sur une plus large gamme que les *microarrays*.

Un autre avantage de la PCR quantitative par rapport aux *microarrays*, c'est l'accès à une mesure absolue de l'expression. En effet, les résultats de *microarrays* sont généralement exprimés en rapports d'activation ou de répression et il est difficile de comparer les résultats d'expression d'une expérience à l'autre.

Le SYBR Green® est plus utilisé que les sondes fluorogènes dans le cas de la validation de résultats de *microarrays* par RT-PCR quantitative car il est plus pratique et moins cher : il n'y a pas besoin de mise au point spécifique ni d'oligonucléotides supplémentaires (Rajeevan *et al.*, 2001). Cependant, même si des plates-formes intégrées à haut débit proposent de coupler *microarrays* et PCR quantitative en plaque 384 puits (Wurmbach *et al.*, 2003), la PCR

quantitative reste un outil de validation d'un petit nombre de gènes qui varie de trois à une cinquantaine selon les articles.

## II. Expression de familles de gènes

### A. Familles de gènes « de maintenance » et expression constitutive

Les gènes dits « de ménage » ou « de maintenance » appartiennent souvent à des familles multigéniques. Ces gènes codent des composants structuraux de la cellule comme les actines ou les tubulines, d'autres appartiennent à des mécanismes fondamentaux tels que la transcription ou la traduction, comme les ARN ribosomiques et les facteurs d'élongation *EFl $\alpha$* , et d'autres encore sont impliqués dans les mécanismes de dégradation des protéines comme les ubiquitines. Nombre de ces gènes dits « de ménage » sont utilisés comme témoins d'expression constitutive lors des études d'expression transcriptionnelle. Sur le plan de l'expression, ces catégories de gènes sont classiquement opposées aux familles de gènes dont la fonction implique une expression spécifique. C'est le cas des gènes du cycle cellulaire, des gènes répondant à différents stimuli comme les attaques de pathogènes, la température ou la lumière, et des gènes impliqués dans un métabolisme particulier, par exemple l'assimilation de l'azote. Une double règle de pensée concernant l'expression des gènes semble se dégager de ces faits : d'une part, les gènes appartenant à des familles dites « de maintenance » présenteraient une expression constitutive et d'autre part, les gènes des autres familles présenteraient une expression spécifique.

Il est cependant important de préciser la notion d'expression constitutive. Le terme constitutif signifie que l'expression d'un gène est similaire d'un échantillon à l'autre et qu'elle peut donc être utilisée comme référence pour normaliser les mesures. Mais, la notion d'expression constitutive varie selon les expérimentations : elle dépend de la nature des échantillons comparés. En effet, lorsque l'expression est mesurée dans une plante soumise à un stimulus, la lumière par exemple, et dans une plante non soumise à ce stimulus, l'expression du gène de référence ne doit pas être modifiée par le stimulus en question. De même, lorsqu'il s'agit de comparer l'expression dans différents organes, il est nécessaire de trouver un gène exprimé avec la même intensité dans tous les organes testés. Ainsi, un gène exprimé spécifiquement dans un organe précis mais non régulé par un stimulus donné répondra à la définition de témoin constitutif dans une étude d'expression en réponse à ce stimulus réalisée uniquement dans cet organe.

Les ARN hélicases à boîte DEAD sont considérées comme une famille de gènes « de maintenance » par leur rôle fondamental dans le métabolisme de l'ARN. L'expression transcriptionnelle constitutive de la majorité des *AtRH* comme celle des facteurs d'élongation

*EF1 $\alpha$*  (Axelos *et al.*, 1989; Liboz *et al.*, 1990) semble conforter l'hypothèse que les familles de gènes « de maintenance » présentent une expression constitutive d'un organe à l'autre. Cependant, les exemples d'étude d'expression transcriptionnelle d'autres familles de gènes « de maintenance » comme les actines (An *et al.*, 1996a; An *et al.*, 1996b; Huang *et al.*, 1996; McDowell *et al.*, 1996a; McDowell *et al.*, 1996b; Huang *et al.*, 1997; Kandasamy *et al.*, 2001), les tubulines  $\alpha$  et  $\beta$  (Ludwig *et al.*, 1987; Oppenheimer *et al.*, 1988; Carpenter *et al.*, 1992; Kim & An, 1992; Kopczak *et al.*, 1992; Carpenter *et al.*, 1993; Chu *et al.*, 1998; Cheng *et al.*, 2001), les ubiquitines (Gausing & Barkardottir, 1986; Callis *et al.*, 1990; Sun & Callis, 1997) et les ATPases-H<sup>+</sup> (Harper *et al.*, 1994; Houlne & Boutry, 1994), montrent au contraire que ces gènes présentent plutôt des profils d'expression organe-spécifiques pouvant se superposer partiellement.

Ces résultats peuvent paraître surprenants car ces gènes sont très souvent utilisés comme contrôle d'expression constitutive. En fait, les différents niveaux et profils d'expression se complètent plus ou moins et, le signal constitutif observé expérimentalement reflète probablement l'abondance des transcrits de l'ensemble de la famille.

Certaines autres familles de gènes, chez lesquelles on pourrait s'attendre à une expression spécifique au vu de leur fonction, présentent une expression similaire dans tous les organes de la plante : les kinases *shaggy AtSK* en sont un exemple (Charrier *et al.*, 2002). Chez les animaux, les villines présentent une expression spécifique, apparemment liée à une fonction spécialisée, alors que chez *A. thaliana*, les *AtVLN* s'expriment dans chaque cellule (Klahre *et al.*, 2000). Les familles de gènes dont l'expression est constitutive sont peu nombreuses dans la littérature, d'autant plus que les expressions spécifiques sont plus souvent recherchées et mises en valeur.

### B. Divergence des profils

Au cours de ce travail de thèse, l'expression de seulement un tiers de la famille a été analysée. L'étude a porté sur les principales branches de l'arbre et sur des gènes dans des situations évolutives différentes. Parmi les *AtRH* étudiés, seulement deux gènes présentent un profil d'expression différent et un niveau d'expression faible : *AtRH58* et *AtRH34*. Le gène *AtRH58* n'a pas d'orthologues dans les autres règnes. La protéine qu'il code est vraisemblablement adressée au chloroplaste. Le gène *AtRH34* présente une plus grande différence d'expression entre les racines et les autres organes que les autres gènes *AtRH*. Sa séquence est très proche de celle d'*AtRH2*. L'orthologue présumé (complémentation non testée) de ces deux gènes

chez *S. cerevisiae* est *FAL1*, qui code une protéine intervenant au cours de la biogenèse des ribosomes. Le faible niveau d'expression et le profil différent des gènes *AtRH58* et *AtRH34* permettent d'envisager un début de divergence de fonction et de considérer l'existence d'un processus de néofonctionalisation pour ces deux gènes par rapport aux autres membres de la famille. Cependant, l'hypothèse d'un transfert du chloroplaste vers le noyau de l'ancêtre du gène *AtRH58* sans modification de sa fonction n'est pas à exclure, auquel cas l'hypothèse de la néofonctionalisation ne serait valable que pour *AtRH34*. D'autres événements évolutifs (néofonctionalisation ou sous-fonctionalisation) pourraient être mis en évidence par une étude complète de la famille.

L'étude de l'expression transcriptionnelle de plusieurs familles de gènes par RT-PCR quantitative a été entreprise chez *A. thaliana*. Même si les protéines de ces familles assurent des fonctions très différentes, allant de la synthèse de la paroi (Yokoyama & Nishitani, 2001) à la signalisation cellulaire (Charrier *et al.*, 2002), en passant par le transport du nitrate (Orsel *et al.*, 2002a) la fermentation alcoolique (Kursteiner *et al.*, 2003) ou la dégradation des protéines (Mladek *et al.*, 2003), ces études partagent certains points communs. Tout d'abord, les familles étudiées sont de taille raisonnable : de 4 à 33 gènes. Ensuite, ces familles sont composées d'une plus ou moins grande proportion de gènes exprimés à des niveaux si faibles qu'il est difficile voire impossible d'étudier leur expression par *Northern blot* (Charrier *et al.*, 2002; Balbi & Lomax, 2003; Kursteiner *et al.*, 2003; Mladek *et al.*, 2003). Au niveau de l'expression, on observe assez fréquemment qu'un ou deux gènes sont fortement exprimés, les autres membres étant limités à des niveaux faibles (Yokoyama & Nishitani, 2001; Kursteiner *et al.*, 2003; Mladek *et al.*, 2003; Tan *et al.*, 2003). Enfin, des profils d'expression spécifiques ne sont observés que pour un nombre réduit de gènes par famille (Yokoyama & Nishitani, 2001; Charrier *et al.*, 2002; Orsel *et al.*, 2002a; Panchuk *et al.*, 2002; Reintanz *et al.*, 2002; Mladek *et al.*, 2003). La différenciation des fonctions au sein des familles multigéniques semble être plus liée à une spécialisation subtile des profils d'expression qu'à une spécialisation drastique. D'autre part, la mise en place de cette spécialisation serait associée à une diminution du niveau de l'expression.

### C. Divergence des niveaux d'expression

#### 1. Proximité des séquences protéiques et niveau d'expression

En comparant les niveaux d'expression des *AtRH* avec leur place dans un arbre de distance basé sur la proximité de leur séquence protéique, on observe qu'au sein d'un groupe de gènes



très proches en séquence, les niveaux ont tendance à être différents (Figure 18, p.103). Comme dix-huit gènes sur vingt présentent le même profil d'expression mais à des niveaux différents, la régulation du niveau d'expression semble être une marque de différenciation des gènes dupliqués. Si un changement de niveau d'expression a une conséquence sur la fonction des gènes, la différence de niveau d'expression pourrait être une marque de divergence de fonction.

## 2. Niveau d'expression et rôle biologique

Chez la tomate, la taille des fruits est majoritairement contrôlée par le QTL *fw2.2* (Frary *et al.*, 2000). La variation de la taille du fruit entre les différentes espèces de tomate est plus associée à la variation du niveau de transcrit *fw2.2* et de sa cinétique d'expression qu'à une variation de la séquence protéique elle-même (Cong *et al.*, 2002; Nesbitt & Tanksley, 2002; Liu *et al.*, 2003). En effet, les protéines FW2.2 des espèces à petits et à gros fruits partagent plus de 98% d'identité (Frary *et al.*, 2000). Comme la taille du fruit est inversement proportionnelle à la quantité de transcrit *fw2.2*, la protéine FW2.2 agirait en exerçant un contrôle négatif sur les divisions cellulaires (Frary *et al.*, 2000). Ainsi, l'exemple du gène *fw2.2* montre que le niveau de transcrit peut avoir, dans certains cas, un effet biologique. Chez *A. thaliana*, il existe au moins 3 orthologues à ce gène : At1g14870, At1g14880 et At5g35525 pour lesquels respectivement 9, 5 et 3 EST sont identifiées.

Une autre indication montrant que les niveaux de transcrit ont une signification biologique réside dans l'obtention de phénotypes différents dans des mutants antisens variant par le niveau d'expression du gène-cible.

Les gènes dont la séquence a peu divergé ont tendance à présenter une expression différente (profil et/ou niveau). Deux hypothèses peuvent expliquer ce fait. D'une part, la pression de sélection sur la partie protéique est différente de celle exercée sur les promoteurs, permettant une variation plus rapide de ces derniers. Les gènes les plus récemment dupliqués, exceptés *AtRH15* et *56*, présentent des répartitions des boîtes télomériques et GGCCA différentes. La divergence des régions promotrices pourrait être un élément précédant le changement de transcription et de fonction.

D'autre part, les mécanismes de duplication ayant conduit à la formation de la famille des *AtRH* font intervenir des rétrotranscriptions (Boudet *et al.*, 2001). Dans ce cas, la partie dupliquée ne concerne que la région exprimée et pas le promoteur, ce qui facilite le

changement de promoteur et l'apparition de nouvelles régulations. Ainsi, plusieurs mécanismes peuvent être impliqués dans la divergence de l'expression des gènes d'une famille et la régulation de l'expression participe au contrôle de la fonction des gènes.

### III. Analyse bioinformatique des promoteurs

Au vu des résultats présentés dans ce manuscrit, le niveau d'expression des gènes est contrôlé au sein d'une famille multigénique. Des éléments intervenant dans cette régulation existent, nous avons repéré notamment l'implication de certains introns et de la boîte TATA.

#### A. Fonctions des boîtes télomériques et GGCCCA

##### 1. Définitions

Chez les eucaryotes, le promoteur d'un gène transcrit par l'ARN polymérase II comprend deux parties. La partie proximale est la zone située autour du site d'initiation de la transcription. Elle contient les éléments pour la mise en place de la machinerie de transcription et le recrutement de l'ARN polymérase II, dont la boîte TATA et le site d'initiation de la transcription. La partie distale, jusqu'à plusieurs milliers de paires de bases en amont, contient notamment les sites de fixation des facteurs de transcription. Classiquement, on implique la région proximale dans le contrôle de l'activité de base du promoteur, c'est-à-dire le contrôle du recrutement des différents facteurs d'initiation de la transcription, et la région distale dans la régulation de l'activité du promoteur (Rombauts *et al.*, 2003). Des exemples montrent que cette vue est simplificatrice et que des éléments présents dans la région proximale peuvent intervenir dans la régulation de l'activité transcriptionnelle (Smale & Kadonaga, 2003). D'autre part, des éléments régulateurs peuvent se trouver en aval du site d'initiation de la transcription, par exemple dans le premier intron du gène (Curie *et al.*, 1993; Norris *et al.*, 1993; Plesse *et al.*, 2001).

D'autres éléments interviennent dans l'activité transcriptionnelle du promoteur, notamment la structure de l'ADN dans l'espace, son état de méthylation ou encore l'enroulement du nucléofilament autour des nucléosomes (Rombauts *et al.*, 2003). En effet, l'activation de la transcription d'un gène commence par une succession de réactions faisant intervenir la phosphorylation puis l'acétylation d'histones et enfin le déplacement du nucléosome qui rendent accessible la boîte TATA et le site d'initiation de la transcription (Featherstone, 2002). Ainsi, les éléments *cis*-régulateurs ne sont pas les seuls éléments contrôlant l'activité d'un promoteur.

##### 2. Les boîtes télomériques et GGCCCA

Nous avons basé la recherche de motifs conservés dans les promoteurs des *AtRH* sur l'hypothèse que, comme ces gènes présentent globalement un profil similaire, ils pouvaient être régulés par les mêmes facteurs et donc présenter des sites de fixation communs. Les motifs

identifiés interviendraient donc sur le profil et non sur le niveau d'expression. Comme aucun lien entre la présence des motifs et l'expression n'a été relevé, on peut proposer l'hypothèse que les boîtes télomériques et GGCCCA ont un autre rôle que celui de recruter des facteurs de transcription, comme par exemple participer à la mise en place de la machinerie de transcription, à la structuration en 3D des promoteurs ou à la méthylation de l'ADN.

La zone proximale des promoteurs est une zone où la déstabilisation des deux brins doit être plus facile pour permettre aisément l'initiation de la transcription. La répartition des insertions de ADN-T dans les collections de mutants d'*A. thaliana* est d'ailleurs préférentielle dans cette zone (Brunaud *et al.*, 2002). La boîte télomérique est également située préférentiellement dans cette région. Ce motif pourrait donc être impliqué dans la facilitation de l'ouverture de la double hélice d'ADN ou dans la protection de l'intégrité de l'ADN au cours de la transcription. Comme les motifs télomériques sont préférentiellement localisés autour du site d'initiation de la transcription, on peut penser qu'ils font parti du corps du promoteur et qu'ils sont impliqués dans la machinerie de l'initiation de la transcription ou de la protection de l'ADN lors de l'ouverture. Cependant, ils ne semblent pas occuper une localisation fixe par rapport au site d'initiation de la transcription : selon le gène la (ou les) boîte(s) télomérique(s) sont situées en aval ou en amont. De plus, il est difficile de justifier la présence préférentielle de ce motif dans le promoteur des *AtRH* par rapport à d'autres catégories de gènes dans le cas d'une fonction liée à la protection de l'intégrité de l'ADN ou de facilitation d'ouverture de la double hélice.

La boîte télomérique pourrait être un site de fixation d'un régulateur négatif de la transcription qui, en se fixant à l'ADN inhiberait l'initiation de la transcription par compétition.

Les boîtes GGCCCA sont toujours situées en amont du site d'initiation de la transcription, ce qui répond mieux à la définition de site de fixation d'un facteur de transcription. La non-mise en évidence de lien entre la présence de cette boîte et l'expression des *AtRH* peut provenir de la combinaison de différents facteurs de transcription au sein des promoteurs. En effet, selon l'hypothèse classique, la régulation de l'expression d'un gène résulte de la combinaison des interactions des différents facteurs de transcription se fixant sur son promoteur. Ainsi, l'action du facteur se liant au motif GGCCCA sera modulée en fonction des autres facteurs régulant l'expression des gènes *AtRH*. Si la composition en sites de fixation de facteurs de transcription diffère entre deux gènes, leur expression variera en conséquence.

### B. D'autres éléments dans les introns

Les gènes *AtRH4*, *19* et *23* sont les orthologues présumés des gènes *ScRH-TIF1* et *-TIF2*, deux gènes fortement transcrits et codant la même protéine, le facteur d'initiation de la traduction eIF4A. Des expériences d'inactivation de gènes indiquent une redondance fonctionnelle complète entre *TIF1* et *TIF2* (Linder & Slonimski, 1989). *AtRH4*, *AtRH19* et *AtRH23* possèdent un intron dans leur région 5' UTR, juste en amont de leur codon d'initiation de la traduction. *AtRH4* et *AtRH19* ont une boîte TATA et sont fortement transcrits alors que le gène *AtRH23* est caractérisé par l'absence d'une boîte TATA et un faible niveau de transcrit. Les arbres de distance établis sur la base des séquences protéiques indiquent que la duplication du gène ancestral à l'origine d'*AtRH23* précède la duplication conduisant aux gènes *AtRH4* et *19*.

Une autre caractéristique des gènes *AtRH4* et *19*, mais absente du gène *AtRH23*, est la présence dans leur intron en 5' UTR d'une séquence codant un petit ARN non-messager nucléaire, les snoR81 et snoR98. SnoR81 et snoR98 sont deux *snoRNA* H/ACA qui guident la pseudo-uridylation de l'ARNr 25S (Marker *et al.*, 2002b) à la même position (*i.e.*  $\psi$  1013). *AtRH4* et *AtRH19* sont les seuls gènes de toute la famille *AtRH* pour lesquels un *snoRNA* a été cloné (groupes H/ACA et C/D) ou prédit (seulement le groupe C/D). Ces deux gènes sont aussi les gènes les plus exprimés des *AtRH*.

Chez les vertébrés, parmi les *snoRNA* localisés dans les introns, un grand nombre appartient à des gènes impliqués dans la biogenèse des ribosomes. Chez *A. thaliana*, un tiers des séquences des *snoRNA* sont localisées dans les introns de gènes codant des protéines ribosomales, des hélicases ou des facteurs d'initiation de la traduction (Marker *et al.*, 2002a). Ces trois catégories « fonctionnelles » de gènes sont celles qui présentent le plus fréquemment une boîte télé (tableau 12, cf partie précédente).

La transcription des *snoRNA* introniques est sous le contrôle du promoteur du gène hôte et de l'épissage du pré-ARNm (Brown *et al.*, 2003a), ce qui implique une expression coordonnée du gène hôte et du *snoRNA*. Une récente hypothèse propose que la conservation de paires de gènes co-régulés, à l'intérieur d'un génome ou entre deux génomes, puisse être une indication d'interaction directe entre les produits de ces gènes co-régulés. Chez les procaryotes et les eucaryotes, des faits s'accumulent en faveur cette hypothèse (van Noort *et al.*, 2003). Chez *A. thaliana*, les deux paires de gènes homologues *AtRH4-snoR81* et *AtRH19-snoR98* sont sous le contrôle du même promoteur et appartiennent tous deux au métabolisme de l'ARN. Pourtant, contrairement aux paires de gènes co-régulés évoquées plus haut, aucune interaction

directe des produits des gènes ne peut être envisagée dans leur cas. En effet, les *snoR81* et *snoR98* sont impliqués dans la modification de l'ARNr 25S. Or, chez la levure *S. cerevisiae*, la maturation des ARNr est un processus complexe qui nécessite l'action de plusieurs ScrH à différentes étapes mais pas celle de TIF1/TIF2, les orthologues potentiels d'*AtRH4* et *AtRH19* (Nissan *et al.*, 2002). De fait, une interaction fonctionnelle entre *AtRH4* et *snoR81* ou *AtRH19* et *snoR98* ne peut être recherchée au cours de la maturation des ribosomes. Ainsi, les *snoR81* et *snoR98* n'interagiraient pas directement avec les hélicases *AtRH4* et *AtRH19*.

Cependant, on peut s'attendre à ce que les ARNm d'*AtRH4* et *AtRH19* interagissent avec l'ARNr 25S modifié par les *snoRNA* *snoR81* et *snoR98* ou avec un ARNr dérivé de l'ARNr 25S modifié. La fonction des modifications nucléotidiques des ARNr par les *snoRNA* est actuellement inconnue. Mais ces modifications sont probablement impliquées dans le contrôle des interactions des ARNr avec d'autres ARN ou protéines (Smith & Steitz, 1997). Dans ce contexte, à la fois la conservation de l'organisation de *AtRH4-snoR81* et *AtRH19-snoR98* et le maintien d'un fort taux de transcription des deux *AtRH* suggèrent que les *snoR81* et *snoR98* auraient un rôle spécifique dans le maintien de forts taux de transcription pour *AtRH4* et *AtRH19* et que la modification spécifique de l'ARNr 25S par les *snoR81* et *snoR98* pourrait augmenter l'efficacité de recrutement des transcrits *AtRH4* et *AtRH19* vers le ribosome ou leur stabilité une fois associés au ribosome. Il en résulterait une augmentation spécifique du taux de transcription réel ou apparent des ces deux gènes.

## Perspectives

Le contexte génomique de la régulation de l'expression des gènes d'*A. thaliana* :  
une approche bioinformatique

### I. Apport de l'approche « analyse d'une famille » à la connaissance du génome

Le séquençage du génome d'*A. thaliana* a révélé l'existence de nombreuses familles multigéniques (AGI, 2000). Vu l'importance de la formation de ces familles dans l'évolution du génome d'*A. thaliana*, l'étude des familles fournit une porte d'entrée de choix dans la compréhension générale de la dynamique du génome. Les structures et les fonctions de plusieurs familles de gènes, dont les ARN hélicases à boîte DEAD (RH), ont été étudiées au laboratoire. Chez *A. thaliana*, les *AtRH* présentent une structure de gènes très variable avec des introns en nombre très différent (0 à 18 selon le gène) et à des positions peu conservées. Le mécanisme de rétrotranscription a joué un rôle important dans l'évolution de la structure des gènes de cette famille (Boudet *et al.*, 2001). La variabilité des structures pourrait participer à l'adaptation de la fonction de ces gènes. L'étude de l'expression d'un tiers des gènes *AtRH* au cours de cette thèse a montré que ces gènes présentent un profil d'expression similaire et des niveaux d'expression très différents. La régulation de l'expression des *AtRH* semble donc très contrôlée.

Des éléments structuraux, comme certains introns ou de courtes séquences (boîtes), sont conservés chez les *AtRH*, laissant supposer un rôle biologique pour ces éléments. A partir des résultats observés à l'échelle d'une famille, il est intéressant, dans une optique d'intégration des données d'expression pour l'analyse fonctionnelle d'un génome, d'étendre l'analyse à l'échelle du génome entier. Jusqu'ici l'analyse était difficile à réaliser pour deux raisons. Premièrement, les annotations de la structure des gènes n'étaient pas de bonne qualité et deuxièmement, des données d'expression homogènes pour l'ensemble du génome étaient peu accessibles et rarement associées aux annotations. A partir du moment où ces données sont rassemblées dans un environnement informatique adéquat, il est possible de réaliser leur intégration et de déterminer des liens entre l'expression et la structure des gènes à l'échelle d'un génome. La base de données FLAGdb<sup>++</sup> a été développée dans le but de mener des analyses intégratives entre l'organisation du génome, la structure des gènes et leur expression. Elle contient actuellement l'ensemble du génome d'*A. thaliana* annoté, associant les données structurales et fonctionnelles des gènes. Cet outil devrait permettre par exemple, de mener une analyse du rôle des introns en 5' UTR dans l'expression des gènes ou d'aborder l'évolution de

la régulation de l'expression des gènes en fonction du type d'évolution des structures des gènes des familles. Les résultats discutés dans cette thèse quant au parallélisme entre le nombre d'EST ou de MPSS et le niveau d'expression des gènes peuvent servir de base pour établir une première approche de l'expression au niveau génomique. L'intégration des données issues des puces à ADN dans la base de données FLAGdb<sup>++</sup> permettrait d'élargir le champ des analyses.

Une autre voie à explorer est la comparaison des données disponibles pour le génome d'*A. thaliana* avec les données établies sur d'autres génomes, comme celui du Riz ou du Maïs. En ajoutant ce genre de données à FLAGdb<sup>++</sup>, il devient possible de mettre en place une approche de génomique comparative et ainsi d'étudier l'implication de l'évolution de l'organisation du génome sur l'expression des gènes et donc sur leur fonction.

## **II. Importance de la topologie dans l'analyse des promoteurs**

Pour étudier la composition des éléments constituant les promoteurs, il est nécessaire de disposer d'un ensemble de gènes dont l'annotation a été expertisée manuellement. Cet ensemble est en cours de réalisation par la participation du laboratoire au projet *GeneFarm*, une base de données contenant l'annotation de gènes d'*A. thaliana* expertisés par famille. Cette méthode garantit une meilleure qualité d'annotation par la connaissance de l'ensemble des paralogues pour réaliser les expertises. Déterminer au mieux le début de la transcription est très important pour pouvoir situer les éléments identifiés dans les promoteurs par rapport à un point fixe. En effet, les boîtes reconnues dans les promoteurs sont petites : quelques paires de bases. Elles peuvent donc facilement apparaître dans la séquence de façon aléatoire. Leur fréquence d'apparition à chaque position par rapport au site d'initiation de la transcription apporte des informations supplémentaires importantes pour distinguer les motifs dus au hasard et ceux répartis de façon non aléatoire pour lesquels peut être supposée une fonction. De plus, l'ordre des éléments identifiés les uns par rapport aux autres renforce également le sens des observations réalisées.

Par ailleurs, une recherche d'éléments de régulation sur l'ensemble du génome donne des résultats bruités. En réalisant des analyses sur des sous-ensembles fonctionnels comme les familles de gènes, les gènes appartenant à une région particulière du génome ou encore l'ensemble des gènes codant les protéines adressées au chloroplaste, les différents types de régulation sont dissociés et le bruitage des données s'en trouve réduit. Ainsi, connaître la



structure intron/exon et la position du site d'initiation de la transcription permet d'établir différentes classes de gènes pour réaliser des analyses différentielles.

### **III. Néo-fonctionnalisation et sous-fonctionnalisation**

A travers l'étude de l'expression d'un tiers des gènes de la famille des ARN hélicases à boîte DEAD, nous avons cherché à établir les mécanismes mis en jeu dans l'évolution de cette famille. La grande variabilité des structures géniques nous laissait supposer une évolution rapide. Cependant, l'acquisition de spécificités transcriptionnelles concerne seulement 10% des gènes de la famille. Ainsi, seule une très faible part de l'augmentation de la taille de la famille par rapport à celle observée dans les génomes animaux est liée à la sous-fonctionnalisation par spécialisation tissulaire. Finalement, le mécanisme de néofonctionnalisation, observé à travers le partage de la transcription entre différents paralogues, serait le principal moteur de l'évolution de la famille des AtRH. A partir d'un gène ancestral, les duplications successives ont conduit à un groupe de gènes, l'un conservant la structure des principaux éléments régulateurs, *i.e.* la boîte TATA et l'intron en 5' UTR, les autres perdant progressivement les deux éléments. A partir de la méthode appliquée au cas particulier de la famille des AtRH se pose maintenant la question de l'impact respectif des mécanismes de néo- et de sous-fonctionnalisation dans l'évolution des fonctions au niveau du génome dans son entier.

## Matériel et Méthodes

### I. Biologie moléculaire

#### A. Matériel biologique

##### 1. Culture

Les graines d'*Arabidopsis thaliana* écotype Columbia 0 (C<sub>0</sub>) sont semées en serre en conditions de jours courts (8 heures d'éclairement, températures jour/nuit : 19.5°C/17.5°C). Au bout de 15 jours, les jeunes plantes sont repiquées en pots de terreau, à raison de 2 plantes par pot et laissées dans les mêmes conditions de culture. Après 3 semaines, les plants sont transférés en conditions de jours longs (16 heures d'éclairement, températures jour/nuit : 19.5°C/17.5°C).

##### 2. Prélèvements

Les différents organes sont prélevés selon le stade de développement des plants d'*A. thaliana*. Les racines, les jeunes feuilles de la rosette (taille inférieure à 1 cm) et les feuilles âgées de la rosette (de 2 à 5 cm) sont récoltées au stade rosette 8 à 12 feuilles ; les feuilles caulines, les tiges inflorescencielles, les très jeunes feuilles autour des bourgeons, les boutons floraux, les fleurs et les jeunes siliques vertes (taille inférieure à 1 cm) après la transition florale. Aussitôt prélevé, chaque échantillon est immédiatement congelé dans l'azote liquide.

#### B. Manipulation d'acides nucléiques

##### 1. Extraction

Les échantillons végétaux sont réduits en fine poudre par broyage au mortier dans l'azote liquide. Ensuite la poudre est répartie en échantillon de 100mg par microtube de 1,5ml ce qui correspond à un volume de poudre d'environ 200µl.

##### a. ADN génomique d'*A. thaliana*

L'ADN génomique est extrait avec le kit DNeasy® plant minikit (Qiagen). L'échantillon est mélangé à un tampon de lyse en présence de ARNase A (1mg/ml) et incubé 10 minutes à 65°C pour dégrader les cellules. Ensuite le détergent, les protéines et les polysaccharides sont précipités par l'ajout d'acide acétique. Après centrifugation, le lysat est filtré sur une colonne permettant d'éliminer la majeure partie des débris. Puis, l'ADN en solution dans le lysat est précipité à l'aide d'éthanol en présence d'hydrochlorure de guanidium (GuHCl). L'échantillon est alors chargé sur une colonne d'affinité pour l'ADN. La colonne est lavée 2

fois avec un tampon contenant de l'éthanol, puis séchée par centrifugation. Enfin, l'ADN est élué dans 50µl d'eau pure après incubation 5 minutes à température ambiante. Une deuxième et une troisième élution sont réalisées afin de récupérer le maximum d'ADN.

#### b. ARN d'*A. thaliana*

Les ARN totaux sont extraits avec le kit RNeasy® plant minikit (Qiagen). Chaque échantillon de 100mg de poudre est solubilisé dans le tampon lyse RLT contenant de l'isothiocyanate de guanidium et du β-mercaptoéthanol (10µl.ml<sup>-1</sup>). Ensuite, l'échantillon est filtré sur une colonne permettant d'éliminer la majeure partie des débris cellulaires. L'ajout d'éthanol absolu permet d'optimiser les conditions de liaison des ARN lorsque le lysat est transféré sur une membrane de silice fixant préférentiellement les ARN. Après un premier lavage, les ARN sont soumis à un traitement à l'ADNase I (3 unités Kunitz par colonne), pendant 30 minutes à température ambiante. Ce traitement permet d'éliminer l'ADN résiduel présent sur la membrane. Il est nécessaire pour des applications très sensibles comme la détection de transcrits faiblement exprimés par PCR quantitative. La colonne est lavée une seconde fois avec le même tampon et deux autres fois avec un deuxième tampon contenant tous les deux de l'éthanol. Après séchage par centrifugation, les ARN sont élués dans 30µl d'eau pure garantie sans ARNase après incubation de 2 minutes à température ambiante. Une deuxième et une troisième élution sont réalisées afin de récupérer le maximum d'ARN. Les échantillons d'ARN ne sont pas précipités avant leur stockage à -80°C.

#### c. ADN plasmidique

Les kits QIAquick spin miniprep® (Qiagen) et Concert™ rapid plasmid purification system (Life Technologies) ont été utilisés pour extraire et purifier l'ADN plasmidique de clones bactériens.

Les bactéries d'une culture de 3ml en phase de croissance exponentielle sont rassemblées au fond d'un microtube par une courte centrifugation. Le milieu de culture est éliminé et le culot de bactéries est resuspendu dans un volume de tampon Tris-HCl 50mM ; EDTA 10mM contenant de l'ARNase A (20mg.ml<sup>-1</sup>). Les cellules sont ensuite lysées par l'ajout d'un volume de NaOH 200mM ; SDS 1% (m/v). La lyse est arrêtée et l'ADN génomique précipité par un tampon neutralisant contenant de l'acide acétique et de l'hydrochlorure de guanidine. Après 10 minutes de centrifugation, le surnageant, contenant l'ADN plasmidique, est récupéré et chargé sur une colonne d'affinité pour l'ADN. L'ADN plasmidique immobilisé sur la

membrane est lavé plusieurs fois en présence d'éthanol et de sels puis élué 2 fois dans 50 $\mu$ l d'eau pure ou de Tris-HCl 10mM ; EDTA 1mM (pH8,0-8,5).

## 2. Précipitation de l'ADN

0,1 volume d'acétate de sodium 3M pH 5,5 sont ajoutés à un volume d'ADN. Le mélange est homogénéisé. L'ADN est laissé à précipiter en présence de 2,5 volumes d'éthanol 100% glacial pendant au moins 30 minutes à  $-20^{\circ}\text{C}$ . La solution est ensuite centrifugée à  $4^{\circ}\text{C}$  pendant 30 minutes. Le surnageant éliminé et remplacé par 2,5 volumes d'éthanol 70% pour laver le culot d'ADN. La solution est centrifugée 15 minutes à  $4^{\circ}\text{C}$  puis le surnageant est éliminé. Les traces d'alcool sont éliminées en plaçant le culot d'ADN sous vide pendant 5 minutes. Enfin, l'ADN est repris dans un volume approprié d'eau pure, de Tris 10mM ou de Tris-HCl 10mM ; EDTA 1mM (pH8,0-8,5).

## 3. Dosage et stockage des acides nucléiques

Les acides nucléiques sont dosés par spectrophotométrie à 260 et 280nm. La concentration des échantillons est estimée par la formule suivante :

$$\text{concentration (ng}\cdot\mu\text{l}^{-1}) = \text{absorbance}_{260\text{nm}} \times R \times \text{facteur de dilution}$$

où R dépend de la nature des acides nucléiques.

ADN double brin :	R = 50
ADN simple brin :	R = 40
ARN simple brin :	R = 37

La pureté des préparations est indiquée par le rapport de l'absorbance à 260nm sur l'absorbance à 280nm. Une préparation d'ADN pure aura un rapport  $\text{abs}_{260\text{nm}}/\text{abs}_{280\text{nm}}$  compris entre 1,7 et 1,9 et une préparation d'ARN entre 1,9 et 2,1.

Les acides nucléiques sont conservés à  $-80^{\circ}\text{C}$  pour une longue durée. Les préparations d'ADN et d'ARN d'utilisation courante sont stockées à  $-20^{\circ}\text{C}$  en fractions aliquotes. Les fractions décongelées utilisées quotidiennement sont conservées à  $4^{\circ}\text{C}$ .

### C. Synthèse d'ADNc simple brin

Les échantillons sont divisés en fractions aliquotes contenant 5 $\mu$ g d'ARN totaux d'*A. thaliana*. Les ARN messagers sont convertis en ADNc simple brin par l'action de la transcriptase inverse Superscript II<sup>TM</sup> en présence d'oligo(dT)<sub>12-18</sub> selon le protocole préconisé par Invitrogen (anciennement Gibco BRL Life Technologies). D'autres enzymes ont été

testées comme l'OmniScript (Qiagen). Aucune différence significative d'efficacité de synthèse n'a été observée. Le choix s'est porté sur le kit SuperScript™ first-strand synthesis system for RT-PCR (Invitrogen) car il permet la conversion d'une quantité maximale d'ARN par unité d'enzyme. Après la synthèse des ADNc simple brin, les ARN sont dégradés par l'action d'une ARNase H. Les ADNc sont ensuite conservés à  $-20^{\circ}\text{C}$ .

#### D. Amplification par PCR

La PCR (ou amplification par réaction de polymérisation en chaîne) est une méthode permettant d'amplifier *in vitro* des fragments d'ADN. Elle nécessite l'utilisation d'une paire d'oligonucléotides capables de s'hybrider sur l'ADN matrice de façon antiparallèle. Ces oligonucléotides sont utilisés comme amorces par une ADN polymérase thermostable qui synthétise le brin complémentaire. Les cycles d'amplification sont composés de trois étapes: la dénaturation de l'ADN matrice, l'hybridation des amorces et la polymérisation à partir des amorces.

##### 1. PCR classique

Le volume réactionnel final de 25  $\mu\text{L}$  contient: 0,7 unité de Taq polymérase Qiagen, son tampon associé, 0,1 mM de chaque oligonucléotide, 50 mM de chaque dNTP (Pharmacia) et la matrice d'ADN. Les réactions ont été effectuées en microtubes, dans un thermocycleur GeneAmp System 9600 de la firme Perkin Elmer Applied Biosystems. Le programme d'amplification est le suivant: 5 min à  $94^{\circ}\text{C}$ , suivies de 40 cycles de 30 s à  $94^{\circ}\text{C}$  puis 1 min à  $55^{\circ}\text{C}$  et 2 min à  $72^{\circ}\text{C}$ . Ensuite le milieu réactionnel est maintenu en attente à  $4^{\circ}\text{C}$ . Plusieurs températures (de 45 à  $62^{\circ}\text{C}$ ) ont été testées pendant l'étape d'hybridation des oligonucléotides, car la température optimale est fonction de leur séquence. La durée de polymérisation (étape à  $72^{\circ}\text{C}$ ) peut varier en fonction de la taille de l'amplifiat attendu. Pour visualiser les résultats, 10  $\mu\text{L}$  sont déposés sur gel d'agarose coloré au BÉt.

Différents types de matrice ont été utilisés en fonction des objectifs de la PCR : pour les réactions effectuées sur de l'ADN génomique ou sur de l'ADN de plasmide purifié, 100 pg à 10 ng d'ADN ont été utilisés comme matrice. Des réactions ont aussi été réalisées directement à partir de colonies de bactéries. Dans ce cas, une infime partie de la colonie est prélevée à l'aide d'un cure-dents et mélangée au milieu réactionnel dans le microtube.

Pour vérifier la présence d'un insert ou pour connaître la taille d'un clone, des oligonucléotides s'hybrident spécifiquement de part et d'autre du site de clonage multiple des vecteurs pUC18 et pBluescript ont été utilisés. Leur séquence est la suivante:

Amorce Directe	5' ACGACGTTGTAAAACGACGGCCAG 3'
Amorce Antisens	5' CAGGAAACAGCTATGACCATGAATTACGC 3'

## 2. PCR quantitative en temps réel

### a. Conditions expérimentales

Les expériences de PCR quantitative sont réalisées en plaques de 96 puits dotées de bouchons prévus pour la détection optique (MicroAmp® Optical 96 Well Reaction Plate and MicroAmp® Optical Caps, Applied Biosystems). Le mélange réactionnel contient dans 25 µL finals, 1,25 unités d'AmpliTaqGold™ dans son tampon associé 1X et 0,01 unité d'EraseUNG, 3 mM de MgCl<sub>2</sub>, 200 mM de chaque dNTP, 100 nM de chaque amorce et la matrice en quantité variable (0,3 à 100 ng d'ADN). Les conditions expérimentales de la PCR quantitative sont similaires à celle de la PCR classique, à la différence près que le milieu réactionnel de PCR quantitative contient, en plus, du SYBR Green®. De plus, la PCR quantitative se déroule en présence de dUTP et non de dTTP. Ainsi l'action de l'EraseUNG™, qui détruit toute molécule d'ADN contenant des dUTP, permet de se prémunir de la contamination des réactions par le produit d'une amplification précédente.

Les gammes-étalon ont été réalisées avec de l'ADN génomique d'*A. thaliana* à 10 ng.ml<sup>-1</sup>. Les échantillons d'ADNc simple brin sont utilisés à raison de 0,01 à 100 ng d'équivalent ARN totaux par réaction.

### b. Cycles thermiques

Le protocole préconisé par Applied Biosystems est une PCR à 2 températures où l'hybridation des amorces et l'élongation se font à 60°C. Le programme est exécuté sur un thermocycleur GeneAmp 9600. Il se compose d'une étape de dénaturation de l'ADN, d'inactivation de l'EraseUNG™ et d'activation de l'AmpliTaqGold™ pendant 10 minutes à 95°C. Ensuite, l'amplification se déroule sur 40 cycles de 15 secondes à 95°C suivies de 1 minute à 60°C. Enfin 10 minutes à 72°C permettent à la polymérase de terminer les amplifications en cours.

### E. Analyse des produits d'amplification : électrophorèse en gel d'agarose

Les fragments d'ADN sont séparés selon leur taille par migration au sein d'un gel d'agarose horizontal baignant dans une solution de TAE 0,5X (Tris-acétate 40 mM ; EDTA 1 mM ; pH 8,0). La concentration en agarose varie entre 0,6% et 2% (poids/volume) suivant la taille des

fragments à séparer ou à visualiser (Sambrook *et al.*, 1989). Le temps et l'intensité de la migration varient en fonction de la taille des fragments à séparer et de la concentration du gel. Pour la majorité des électrophorèses réalisées, une cuve de type Mupid-2 (Cosmo Bio Co.) a été utilisée (TAE 0,5X, migration de 20 min à 50 V). Après migration, le gel est baigné dans une solution de bromure d'éthidium ( $1 \text{ mg.ml}^{-1}$ ) pendant 10 min sous agitation lente puis rincé à l'eau distillée. Le gel éclairé aux UV est en suite photographié par une caméra CCD. Les marqueurs de taille déposés sont, soit le "1 kb ladder" (Gibco BRL) pour estimer la taille d'un fragment, soit l'ADN du phage Lambda digéré par HindIII, afin de quantifier l'ADN déposé sur le gel. Les échantillons sont chargés dans un tampon coloré et dense (Bleu de Bromophénol 0,4% (p/v) ; EDTA 50 mM ; glycérol 50% (v/v)).

## F. *Dot Blot*

### 1. Principe

Le principe de la technique de *dot blot* est l'hybridation d'une sonde nucléique sur de l'ADN (vecteur, produit PCR ou ADNc) fixée sur une membrane de nylon. Elle tire son nom de la manière dont l'ADN est déposé sur la membrane : chaque échantillon apparaît sous la forme d'un disque de quelques millimètres de diamètre, donnant l'impression de points répartis sur la membrane.

Le dépôt se fait par l'intermédiaire d'un appareil à vide qui permet de déposer simultanément 96 échantillons sur la membrane. Les échantillons d'ADN sont dénaturés 5 à 10 min à  $100^{\circ}\text{C}$ , puis refroidis brutalement sur de la glace. Ils sont déposés par fraction de  $50\mu\text{l}$  par puits. Ensuite, les puits sont lavés avec  $100\mu\text{l}$  d'eau pure. Enfin l'ADN est fixé à la membrane en l'exposant à un rayonnement U.V (dose =  $1400 \text{ joules/m}^2$ ).

### 2. Hybridation

La membrane est préhybridée dans un tampon 5X SSC (NaCl 0,75 M ; citrate de sodium 75 mM; pH7); 5X Denhardt (ficol 0,1% v/v, polyvinylpyrrolidone 0,1% p/v, sérum albumine bovine 0,1% p/v) ; 0,5% SDS pendant au minimum 1 heure à  $65^{\circ}\text{C}$ . Elle est ensuite incubée à  $65^{\circ}\text{C}$  dans un tampon 5X SSC ; 5X Denhardt ; 0,5% SDS en présence de la sonde radioactive préalablement dénaturée pendant toute la nuit. La radioactivité initiale de la solution est de  $2.10^6 \text{ cpm.ml}^{-1}$ . Le tampon d'hybridation contenant la sonde radioactive est éliminé et la membrane est lavée dans plusieurs tampons dont la stringence augmente au fur et à mesure pour éliminer au maximum l'hybridation aspécifique de la sonde. Chaque lavage est effectué pendant 15 minutes à  $65^{\circ}\text{C}$ . Les 2 premiers lavages se déroulent dans un tampon 2X SSC;

0,1% SDS, puis les 2 lavages suivants dans un tampon 1X SSC; 0,1% SDS. Le cinquième lavage se fait dans un tampon 0,4X SSC; 0,1% SDS et le dernier lavage dans un tampon 0,1X SSC; 0,1% SDS.

La membrane est emballée dans un film scellé et exposée à un film radiographique ou à un écran pour phosphorImager®. L'image de la membrane est obtenue par révélation du film radiographique. L'analyse de l'écran par le phosphorImager® permet de quantifier les signaux de la membrane.

### 3. Purification des sondes radiomarquées

Les sondes sont radioactivement marquées à l'aide de [ $\alpha$ -<sup>32</sup>P]dCTP selon le protocole décrit par Amersham (Ready to go™ DNA labelling beads). 50 ng de produit PCR sont dénaturés par la chaleur pendant 5 à 10 minutes puis mis en présence de radionucléotides et de klenow (ADN polymerase). Après synthèse (1 heure à 37°C) la sonde est séparée des radionucléotides non incorporés par chromatographie d'exclusion sur une colonne de résine Sephadex G-50. La radioactivité des fractions est mesurée dans un compteur à scintillation et les fractions contenant la sonde sont rassemblées.

### 4. Préparation de la résine Sephadex G-50 pour la purification des sondes radioactives par chromatographie par exclusion

10g de résine Sephadex G-50 sont suspendus dans un volume final de 160 ml d'eau pure. La résine est vigoureusement agitée puis laissée à décanter ou centrifugée 1 minute à 4000 rpm. L'eau et les billes qui flottent à la surface sont éliminées. L'opération est répétée 2 fois en complétant à chaque fois le volume à 160 ml avec de l'eau pure. La résine est ensuite équilibrée avec une solution de Tris-HCl 10mM ; EDTA 1mM (pH 7,6). La résine est autoclavée 15 minutes à 110-120°C dans une bouteille en verre, puis elle est répartie par fractions de 40 ml dans des tubes de 50ml. Après décantation, le surplus de TE est éliminé et remplacé par le volume approprié de Tris-HCl 10mM ; EDTA 1mM ; NaCl 100mM (pH8) stérile (volume final = 40ml). La résine est remise en suspension par une agitation vigoureuse. Après décantation, le surnageant est éliminé. La résine est équilibrée dans le même volume de TEN stérile et stockée à température ambiante.

### G. Séquences

Les produits PCR sont tout d'abord purifiés sur colonne avec le kit « QIAquick PCR Purification » (Qiagen). Puis leur concentration est évaluée sur gel. Ensuite les réactions de



séquences proprement dites sont réalisées en chimie Dye Terminator avec le kit « Big Dye™ Terminator » (Applied Biosystems) sur 30 ng de produits PCR selon la technique de Sanger et précipitées conformément aux instructions du fabricant. Le protocole préconisé est suivi à la lettre. Les réactions sont réalisées avec un thermocycleur GeneAmp System 9600 (Perkin Elmer Applied Biosystems) et les séquences sont obtenues et analysées sur un séquenceur automatique ABI 373 DNA Sequencer avec stretch (Applied Biosystems).

## II. Bioinformatique

### A. Analyse de séquences

#### 1. Recherche de similarités

*BLAST* (Altschul *et al.*, 1997) est un logiciel de recherche de similitudes de séquences dans différentes banques de séquences nucléiques et protéiques. Pour les séquences nucléiques, il attribue un score positif aux appariements et un score négatif aux mésappariements. Pour les séquences protéiques, il utilise une matrice de scores de similitude entre deux résidus (BLOSUM, PAM), attribuant des scores positifs aux identités et aux remplacements conservatifs et des scores négatifs aux remplacements non conservatifs. En comparant localement deux séquences, il recherche le score de similitude le plus élevé sur une fenêtre de taille définie. Il essaie ensuite d'élargir l'alignement, jusqu'à atteindre la fin de la séquence, un score de zéro ou que le score ait diminué d'une certaine quantité. Le résultat de la recherche comprend une liste des séquences similaires (numéros d'accèsion et en-têtes des annotations), et des alignements de séquences correspondants. Les différents algorithmes de *BLAST* pour les recherches de similitude sont donnés dans le tableau ci-dessous.

Programme	Séquence ( <i>Query</i> )	Banque ( <i>Subject</i> )
<b>BLASTn</b>	séquence nucléotidique	séquences nucléotidiques
<b>BLASTx</b>	séquence nucléotidique traduite	séquences protéiques
<b>BLASTp</b>	séquence protéique	séquences protéiques
<b>tBLASTn</b>	séquence protéique	séquences nucléotidiques traduites
<b>tBLASTx</b>	séquence nucléotidique traduite	séquences nucléotidiques traduites

NB : lorsqu'une séquence nucléotidique est traduite, les six phases de lecture sont considérées.

#### 2. Alignement de séquences

##### a. Alignement multiple

Le logiciel *ClustalX* (Thompson *et al.*, 1997) permet d'aligner simultanément plusieurs séquences. La méthode consiste à aligner deux à deux les séquences, pour pouvoir calculer une matrice de distances estimant la divergence entre chaque séquence. Les taux de similitude et d'identité sont calculés sur toute la longueur des séquences. L'introduction d'espaces dans l'alignement est prise en compte. Un arbre des distances peut ensuite être construit à partir de cette matrice (Thompson *et al.*, 1994). J'ai utilisé ce logiciel pour déterminer les zones divergentes entre des gènes proches lors de la conception des amorces de PCR afin d'assurer un maximum de spécificité.

### b. Alignement de transcrits

*Sim4* est un logiciel spécifiquement conçu pour aligner des séquences de transcrits avec une séquence génomique (Florea *et al.*, 1998). Basé sur une méthode de type BLAST, il a la particularité, d'une part, de tester l'alignement de la séquence du transcrit dans les deux sens par rapport à la séquence génomique, et d'autre part, de couper préférentiellement les alignements aux sites d'épissage (GT-AG et CT-AC). Ce logiciel n'aligne qu'un transcrit à la fois. Lorsque plusieurs transcrits sont disponibles pour un même gène, un programme développé au laboratoire (*mSim2GDE.pl*) permet d'aligner plusieurs transcrits successivement avec la même séquence génomique et de transférer les résultats de *Sim4* dans un fichier contenant un alignement multiple lisible par GDE (*Genetic Data Environment*). J'ai utilisé ce logiciel pour annoter la structure des gènes *AtRH* afin de déterminer notamment leur région 5' UTR.

### 3. Outil de visualisation

Installé sur une station de travail Sun, GDE (Smith *et al.*, 1994) propose une interface graphique permettant, entre autres, d'éditer des alignements de séquences afin de définir par exemple la structure intron/exon des gènes. Il est aussi utile pour vérifier les données de séquençage. Ce logiciel peut lire plusieurs types de fichiers de séquences. Le plus simple est un fichier texte présentant chaque séquence précédée d'une ligne commençant par le caractère # pour une séquence d'ADN et % pour une séquence protéique, suivi éventuellement par le nom de la séquence. Il est également capable d'interpréter les fichiers au format GENBANK. Enfin, il génère des fichiers à son propre format. Dans ce format de fichier dit .gde, chaque séquence est associée à un certain nombre d'informations comme la nature de la séquence (ADN ou protéine), sa position dans l'alignement, son nom et des commentaires. Toutes ces informations, y compris la séquence, sont déterminées par des mots-clé dans un bloc défini par des accolades. La succession des blocs détermine l'ordre des séquences apparaissant à l'écran. Tout le travail d'annotation structurale des gènes *AtRH* a été réalisé avec cette interface.

### B. Outils informatiques pour le choix des couples d'amorces

La qualité de la PCR dépend principalement des amorces. Elles doivent être spécifiques et permettre une amplification efficace. La spécificité est recherchée en plaçant les amorces dans les parties les plus divergentes des gènes. L'efficacité de la PCR est maximisée en sélectionnant les amorces dont la probabilité de former de structures secondaires et/ou de

dimères entre elles est minimale. Les amorces sont sélectionnées à l'aide des logiciels suivants : *Oligo 4.0*, *Primer Express*<sup>TM</sup> et *Amplify 1.2*.

### 1. Oligo 4.0

Le logiciel *Oligo 4.0* constitue le point central de la sélection des amorces. Il permet d'évaluer la capacité des amorces à s'associer en dimères et/ou à former des structures secondaires de type épingle à cheveux. La stabilité des appariements est fonction du calcul du  $\Delta G$  (Rychlik & Rhoads, 1989). Plus le  $\Delta G$  est négatif, plus les appariements sont stables. Pour éviter les dimères, le  $\Delta G$  des appariements doit être supérieur à -10 kcal/mol. Il faut surtout limiter la formation de dimères avec des extrémités 3' sortantes, qui permettent l'amorçage d'une amplification non désirée. Pour éviter les boucles trop stables, le  $\Delta G$  doit être positif.

### 2. Primer Express<sup>TM</sup>

Ce logiciel est préconisé par Applied Biosystems pour définir des couples d'amorces qui respectent les conditions expérimentales de la PCR quantitative. Je l'ai utilisé uniquement pour vérifier la  $T_m$ , qui doit être comprise entre 58 et 60°C, et la taille de l'amplifiat (50 à 200 pb). Primer Express utilise la méthode du plus proche voisin (*nearest-neighbor method*) pour calculer la  $T_m$  des oligonucléotides (Breslauer *et al.*, 1986).

### 3. Amplify 1.2

Le logiciel *Amplify 1.2* simule une expérience de PCR (Engels, 1993). Il analyse la possibilité d'hybridation des amorces sur une séquence-cible (ou matrice) et détermine les portions de la séquence-cible potentiellement amplifiées. Il détermine la stabilité des amorces et leur force d'interaction avec la matrice. Il apporte aussi des informations complémentaires sur les produits amplifiés (longueur, composition, probabilité d'amplification), les sites de liaisons des amorces et la possibilité de dimérisation des amorces. Pour de la conception des amorces de PCR quantitative, la matrice utilisée est constituée de l'ensemble des séquences connues d'ADNc de la famille *AtRH*. Ainsi les couples les plus spécifiques, c'est-à-dire ceux qui présentent le risque minimum d'amplifier un fragment sur un autre ADNc que celui pour lequel ils ont été choisis, ont été retenus.

## C. Traitements des données de PCR quantitative

### 1. GeneAmp 5700SDS

Le logiciel d'analyse *GeneAmp5700SDS* (Applied Biosystems) stocke les données des mesures obtenues par PCR quantitative dans un fichier au format .csv (*coma separated values*). Ainsi, une ligne contient les données relatives à un puits de la plaque sous forme de valeurs séparées par des virgules. Ce type de fichier est reconnu par le logiciel Excel (Microsoft) qui l'interprète sous forme d'un tableur, les valeurs d'une mesure apparaissant dans les cases successives d'une ligne. Grâce à une possibilité fonctionnelle d'Excel (les macros), j'ai pu automatiser l'analyse de données brutes. Cette automatisation permet notamment de modifier l'affichage du tableau initial et de réaliser plusieurs calculs. Ce fichier interprété est enregistré au format .xls (feuille de calcul Microsoft Excel).

### 2. Macros Excel

L'automatisation du traitement des données issues de la PCR quantitative a été réalisée en langage VBA (Visual Basic for Application) par modification de macros enregistrées. L'enregistrement d'une macro est l'un des outils proposés par Excel. Avec cet outil, le logiciel génère automatiquement le code d'un programme en enregistrant la suite d'actions réalisées par l'utilisateur, par exemple écrire « AtRH4 » dans la case A2, puis sélectionner une plage de cellule et la colorer en gris. A chaque fois que l'utilisateur lance l'exécution de cette macro, l'ordinateur répète les actions et écrit «AtRH4» dans la case A2 puis colore en gris la plage de cellule déterminée par le programme.

Le code de la macro est accessible et modifiable dans l'éditeur de code VBA de l'environnement Microsoft. Ce code « spécifique » peut être transformé en un code plus générique en y introduisant d'autres instructions comme des conditions d'exécution ou des boucles. Ainsi, les instructions précédentes peuvent être modifiées de façon à ce que le code soit exécuté pour différentes cases. On obtient alors le code correspondant à l'algorithme suivant :

```
Ecrire le contenu d'une variable dans la première case de la ligne  
Faire jusqu'à la fin du tableau [  
    Si la cellule active contient « STND », alors  
        colorer les 4 cellules à sa droite en gris,  
        puis passer à la ligne suivante  
    Sinon traiter la ligne suivante.  
]
```

De plus, l'environnement Microsoft donne la possibilité de personnaliser des boîtes de dialogue avec l'utilisateur. On pourra donc modifier le programme de telle sorte qu'il ouvre

une fenêtre demandant de saisir le nom du gène étudié à l'utilisateur (par exemple AtRH4) avant de l'écrire dans les cases voulues et de colorer d'autres cases en gris.

Un « nettoyage » du code est en général nécessaire. En effet, l'ensemble des paramètres associés à l'objet modifié est repris dans le code, alors que l'action de l'utilisateur ne modifie la valeur que d'un seul paramètre à la fois. Cela engendre beaucoup de lignes inutiles car les valeurs par défaut des paramètres non modifiés n'ont pas besoin d'être précisées dans le code.

L'intérêt des macros enregistrées réside donc dans la modification de leur code permettant l'adaptation des tâches exécutées aux données fournies par l'utilisateur. Cependant, le code résultant de l'enregistrement d'une macro est dicté par l'enchaînement séquentiel des actions. Si un groupe d'actions se répète à différents intervalles, le code correspondant sera écrit plusieurs fois à l'identique aux différents endroits du programme. Ce type de programmation est peu élégant d'un point de vue informatique et coûteux en mémoire mais il a le mérite d'être simple. Ces macros pourraient donc être optimisées en modifiant le code de façon à supprimer les parties de code répétées.

Pour l'automatisation de l'analyse des résultats de PCR quantitative j'ai créé 8 macros. La première permet de transformer les données brutes d'une expérience de PCR quantitative en données interprétées. Elle contient à la fois des instructions de mise en page et des calculs. Les données sont séparées en autant de couples de feuilles que de gènes mesurés. Pour chaque gène, la première feuille nommée « Data » contient tous les résultats de l'expérience de PCR quantitative : nom et type de l'échantillon, couple d'amorces utilisé, Ct et écart-type des Ct des échantillons identiques, nombre de copies du gène, moyenne et écart-type du nombre de copies des échantillons identiques pour chaque mesure d'une part et, le seuil de fluorescence, la pente de la gamme-étalon et l'efficacité de la PCR d'autre part. Les échantillons dont le Ct est hors de la gamme sont indiqués en italique et ne sont pas reportés sur la deuxième feuille. La seconde feuille nommée « Bilan » contient l'analyse des données quantitatives relatives aux échantillons analysés : quantité brute de transcrits, quantité par ng d'ARN totaux, moyenne et écart-type, et un tableau récapitulatif.

La seconde macro permet d'extraire gène par gène les données de chaque expérience, de les enregistrer dans des fichiers séparés et de ranger ces fichiers dans les répertoires associés aux gènes étudiés.

Les suivantes permettent de récapituler et synthétiser les données extraites des différents fichiers, pour réaliser notamment des moyennes de plusieurs expériences indépendantes.

## D. Analyse des promoteurs

### 1. Séquences

Les séquences utilisées pour les analyses de promoteurs correspondent aux 1000 pb situées en amont de l'ATG initiateur des 56 gènes *AtRH*. La région 5'UTR des gènes est donc incluse dans les analyses. Les deux pseudogènes de cette famille ont été écartés de l'étude. La position de l'ATG a été déterminée par comparaison de la séquence génomique avec les séquences d'ARNm complets ou partiels (EST) disponibles dans les bases de données. Lorsque aucune séquence d'ARNm n'était disponible, l'ATG initiateur a été déterminé sur la base de prédictions informatiques (NetGene2 et GeneMarkHMM), en tenant compte de la structure des gènes proches.

Les séquences promotrices ont été rassemblées dans un même fichier au format fasta, c'est-à-dire que les séquences sont présentées les unes après les autres, chaque séquence étant précédée d'une ligne commençant par le caractère > suivi du nom du gène.

### 2. Analyse de courbure

La séquence promotrice des 56 *AtRH* a été analysée avec des algorithmes associés au logiciel *ADN-viewer* du LIMSI (Gherbi & Herisson, 2002). *ADN-viewer* permet de calculer, visualiser et manipuler (zoom, rotation, extraction d'une portion, choix de la représentation) la représentation en 3D de la trajectoire que prendrait une séquence d'ADN nu sous la seule contrainte de l'enchaînement de ses bases. Cette représentation en 3D des séquences est calculée à partir du modèle proposé par Bolshoy et Trifonov définissant les angles formés par deux plateaux de bases successifs (Bolshoy *et al.*, 1991).

A partir de la séquence textuelle, la trajectoire de chaque séquence est calculée dans un espace à 3 dimensions et une projection selon les 3 axes est obtenue. La courbure est ensuite calculée à partir des fichiers de projections par la méthode de calcul des « abscisses curvilignes » (le calcul prend en compte l'angle et la distance de la trajectoire), en précisant les paramètres suivants : un pas de déplacement dans la séquence correspondant à 1 pb et une granularité (taille de la fenêtre pour le calcul) correspondant à 40 pb. L'algorithme calcule la courbure sur une fenêtre de séquence puis déplace cette fenêtre le long de la séquence.

La représentation 3D de la trajectoire, les projections et la courbure sont ensuite visualisées à l'aide de l'outil *Scilab* (Gomez, 1999).

### 3. Recherche de motifs dans les promoteurs

#### a. PlantCARE

La base de données PlantCARE<sup>21</sup> contient l'ensemble des motifs *cis*-régulateurs identifiés chez les végétaux et propose des outils pour l'analyse informatique de séquences promotrices (Lescot *et al.*, 2002). L'outil « *Search for CARE* » permet de rechercher dans une séquence la présence des motifs répertoriés dans la base. Cet outil a été utilisé pour déterminer la présence de boîte TATA dans la séquence des promoteurs des *AtRH*. Nous avons retenu au plus une seule boîte TATA par gène, située à 30 nt environ du site apparent du début de la transcription (moyenne : 38 nt, min : 21 nt, max : 85 nt).

L'outil « *Search for CARE* » a également été utilisé pour rechercher des motifs de régulation par la lumière dans le promoteur du gène *AtRH58*.

#### b. RSAT, outil d'analyse de séquences régulatrices

Le logiciel RSAT<sup>22</sup> (*Regulatory Sequences Analysis Tool*) permet d'identifier des motifs communs à un ensemble de séquences issues d'un même organisme, sur la base d'une analyse statistique de la composition des séquences (van Helden *et al.*, 1998). Pour ce faire, le logiciel décompose les séquences en oligomères, et calcule la fréquence de chaque oligomère. Il compare ensuite la fréquence obtenue avec la fréquence attendue de chaque oligomère dans le génome dont sont issues les séquences analysées. Les oligomères sur-représentés, c'est-à-dire présentant une fréquence plus importante que la fréquence attendue, sont retenus puis assemblés pour composer des motifs plus longs.

Le fichier contenant les 56 séquences promotrices au format fasta a été analysé avec l'outil *oligo-analysis* qui détermine la fréquence des oligomères. Deux paramètres ont été précisés : l'organisme de référence (*A. thaliana*) et le modèle de fréquences attendues (ou *background*). Le *background* sélectionné est « *upstream* », c'est-à-dire l'ensemble des régions 5' UTR des gènes prédits par AGI (séquences de 2000 pb en amont du premier ATG de chaque gène). Les autres possibilités de *background* sont : « *upstream no ORF* » qui correspond aux séquences précédentes sans chevauchement avec la séquence d'un éventuel gène en amont et « *intergenic* » qui correspond à l'ensemble des séquences hors gènes, régions 5' UTR et 3' UTR incluses. Le choix du modèle de référence est fonction de la nature des séquences analysées. Les séquences de *background* doivent présenter des caractéristiques identiques aux

---

<sup>21</sup> PlantCARE <http://intra.psb.ugent.be:8080/PlantCARE/>

<sup>22</sup> RSAT <http://rsat.ulb.ac.be/rsat/>



séquences analysées ou être les plus approchantes. Dans notre étude, seule la taille des séquences à analyser et celles de référence diffèrent : 1000 pb pour les promoteurs des *AtRH* et 2000 pb pour le *background*. Les autres paramètres du logiciel ont été laissés par défaut : oligomères de 6 nt, séquences au format fasta, filtrage des séquences et analyse des deux brins.

*Oligo-analysis* génère un tableau contenant les hexamères sur-représentés accompagnés de leurs caractéristiques statistiques. Les hexamères sont classés en fonction de leur indice de significativité. Plus l'indice est grand, plus l'hexamère correspondant est sur-représenté dans le jeu de séquences analysé par rapport au jeu de séquences de référence. En pratique, seuls les hexamères dont l'indice est supérieur à 1 sont retenus.

Les données générées par *oligo-analysis* sont directement transmises à l'outil *pattern-assembly* qui aligne les hexamères et détermine des consensus de motifs plus longs. Les alignements sont ensuite inspectés visuellement pour valider les motifs proposés.

Les positions de ces motifs dans les séquences promotrices des 56 *AtRH* ont été déterminées à l'aide d'un programme développé au laboratoire : *motifSearch.pl*.

## Bibliographie

- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadiou E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*. 287: 2185-2195.
- Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C & Venter JC (1992) Sequence identification of 2,375 human brain genes. *Nature*. 355: 632-634.
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF & *al.* (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*. 252: 1651-1656.
- Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O & *al.* (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*. 377: 3-174.
- Adams MD, Soares MB, Kerlavage AR, Fields C & Venter JC (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat Genet*. 4: 373-380.
- AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 408: 796-815.
- Aharoni A & O'Connell A (2002) Gene expression analysis of strawberry achene and receptacle maturation using DNA microarrays. *J Exp Bot*. 53: 2073-2087.
- Ahmed FE (2002) Detection of genetically modified organisms in foods. *Trends Biotechnol*. 20: 215-223.
- Akimoto-Tomiyama C, Sakata K, Yazaki J, Nakamura K, Fujii F, Shimbo K, Yamamoto K, Sasaki T, Kishimoto N, Kikuchi S, Shibuya N & Minami E (2003) Rice gene expression in response to N-acetylchitooligosaccharide elicitor: comprehensive analysis by DNA microarray with randomly selected ESTs. *Plant Mol Biol*. 52: 537-551.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25: 3389-3402.
- An YQ, Huang S, McDowell JM, McKinney EC & Meagher RB (1996a) Conserved expression of the *Arabidopsis* ACT1 and ACT 3 actin subclass in organ primordia and mature pollen. *Plant Cell*. 8: 15-30.
- An YQ, McDowell JM, Huang S, McKinney EC, Chambliss S & Meagher RB (1996b) Strong, constitutive expression of the *Arabidopsis* ACT2/ACT8 actin subclass in vegetative tissues. *Plant J*. 10: 107-121.
- Anantharaman V, Koonin EV & Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res*. 30: 1427-1464.
- Angelastro JM, Klimaschewski LP & Vitolo OV (2000) Improved NlaIII digestion of PAGE-purified 102 bp ditags by addition of a single purification step in both the SAGE and microSAGE protocols. *Nucleic Acids Res*. 28: E62.
- Anterola AM, Jeon JH, Davin LB & Lewis NG (2002) Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *J Biol Chem*. 277: 18272-18280.
- Arguello-Astorga GR & Herrera-Estrella LR (1996) Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. *Plant Physiol*. 112: 1151-1166.
- Arumuganathan K & Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep*. 9: 208-218.
- Asamizu E, Nakamura Y, Sato S & Tabata S (2000) A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res*. 7: 175-180.

- Asano T, Masumura T, Kusano H, Kikuchi S, Kurita A, Shimada H & Kadowaki K** (2002) Construction of a specialized cDNA library from plant cells isolated by laser capture microdissection: toward comprehensive analysis of the genes expressed in the rice phloem. *Plant J.* 32: 401-408.
- Aubourg S** (1999) Les familles multigéniques comme étape dans l'analyse fonctionnelle des gènes : quelques exemples tirés de l'étude de séquences génomiques d'*Arabidopsis thaliana*. Thèse Paris XI. Orsay.
- Aubourg S, Boudet N, Kreis M & Lecharny A** (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. *Plant Mol Biol.* 42: 603-613.
- Aubourg S, Kreis M & Lecharny A** (1999) The DEAD box RNA helicase family in *Arabidopsis thaliana*. *Nucleic Acids Res.* 27: 628-636.
- Aubourg S, Takvorian A, Cheron A, Kreis M & Lecharny A** (1997) Structure, organization and putative function of the genes identified within a 23.9-kb fragment from *Arabidopsis thaliana* chromosome IV. *Gene.* 199: 241-253.
- Audic S & Claverie JM** (1997) The significance of digital gene expression profiles. *Genome Res.* 7: 986-995.
- Axelos M, Bardet C, Liboz T, Le Van Thai A, Curie C & Lescure B** (1989) The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 alpha: molecular cloning, characterization and expression. *Mol Gen Genet.* 219: 106-112.
- Bachelier JP & Cavaille J** (1997) Guiding ribose methylation of rRNA. *Trends Biochem Sci.* 22: 257-261.
- Balbi V & Lomax TL** (2003) Regulation of early tomato fruit development by the *Diageotropica* gene. *Plant Physiol.* 131: 186-197.
- Bancroft I** (2001) Duplicate and diverge: the evolution of plant genome microstructure. *Trends Genet.* 17: 89-93.
- Barakat A, Szick-Miranda K, Chang IF, Guyot R, Blanc G, Cooke R, Delseny M & Bailey-Serres J** (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* 127: 398-415.
- Barry CS, Blume B, Bouzayen M, Cooper W, Hamilton AJ & Grierson D** (1996) Differential expression of the 1-aminocyclopropane-1-carboxylate oxidase gene family of tomato. *Plant J.* 9: 525-535.
- Baumgarten A, Cannon S, Spangler R & May G** (2003) Genome-level evolution of resistance genes in *Arabidopsis thaliana*. *Genetics.* 165: 309-319.
- Becker JD, Boavida LC, Carneiro J, Haury M & Feijo JA** (2003) Transcriptional profiling of *Arabidopsis* tissues reveals the unique characteristics of the pollen transcriptome. *Plant Physiol.* 133: 713-725.
- Becker-Andre M & Hahlbrock K** (1989) Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res.* 17: 9437-9446.
- Bennetzen JL** (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol.* 42: 251-269.
- Bennetzen JL** (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica.* 115: 29-36.
- Berger S, Mitchell Olds T & Stotz HU** (2002) Local and differential control of vegetative storage protein expression in response to herbivore damage in *Arabidopsis thaliana*. *Physiol Plant.* 114: 85-91.
- Blanc G, Barakat A, Guyot R, Cooke R & Delseny M** (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell.* 12: 1093-1101.
- Bolle C, Herrmann RG & Oelmüller R** (1996) Intron sequences are involved in the plastid- and light-dependent expression of the spinach Psad gene. *Plant J.* 10: 919-924.
- Bolshoy A, McNamara P, Harrington RE & Trifonov EN** (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc Natl Acad Sci U S A.* 88: 2312-2316.
- Bonaldo MF, Lennon G & Soares MB** (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* 6: 791-806.
- Bonaventure G, Salas JJ, Pollard MR & Ohlrogge JB** (2003) Disruption of the FATB gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell.* 15: 1020-1033.
- Bond AT, Mangus DA, He F & Jacobson A** (2001) Absence of Dbp2p alters both nonsense-mediated mRNA decay and rRNA processing. *Mol Cell Biol.* 21: 7366-7379.
- Boudet N** (2002) Analyse de l'évolution du génome d'*Arabidopsis thaliana* par l'étude des familles de gènes. Thèse Paris XI. Orsay.
- Boudet N, Aubourg S, Toffano-Nioche C, Kreis M & Lecharny A** (2001) Evolution of intron/exon structure of DEAD helicase family genes in *Arabidopsis*, *Caenorhabditis*, and *Drosophila*. *Genome Res.* 11: 2101-2114.
- Brander KA & Kuhlemeier C** (1995) A pollen-specific DEAD-box protein related to translation initiation factor eIF-4A from tobacco. *Plant Mol Biol.* 27: 637-649.
- Brander KA, Mandel T, Owtrim GW & Kuhlemeier C** (1995) Highly conserved genes coding for eukaryotic translation initiation factor eIF-4A of tobacco have specific alterations in functional motifs. *Biochim Biophys Acta.* 1261: 442-444.

- Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca Serra P & Sansone SA** (2003) ArrayExpress - a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 31: 68-71.
- Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J & Corcoran K** (2000a) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol.* 18: 630-634.
- Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C, Mao JI, Luo S, Kirchner JJ, Eletr S, DuBridge RB, Burcham T & Albrecht G** (2000b) In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci U S A.* 97: 1665-1670.
- Breslauer KJ, Frank R, Blocker H & Marky LA** (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A.* 83: 3746-3750.
- Breyne P, Dreesen R, Cannoot B, Rombaut D, Vandepoele K, Rombauts S, Vanderhaeghen R, Inze D & Zabeau M** (2003) Quantitative cDNA-AFLP analysis for genome-wide expression studies. *Mol Gen Genet.* 269: 173-179.
- Brodmann PD, Ilg EC, Berthoud H & Herrmann A** (2002) Real-time quantitative polymerase chain reaction methods for four genetically modified maize varieties and maize DNA content in food. *J AOAC Int.* 85: 646-653.
- Brown JW, Clark GP, Leader DJ, Simpson CG & Lowe T** (2001) Multiple snoRNA gene clusters from *Arabidopsis*. *RNA.* 7: 1817-1832.
- Brown JW, Echeverria M & Qu LH** (2003a) Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci.* 8: 42-49.
- Brown RL, Kazan K, McGrath KC, Maclean DJ & Manners JM** (2003b) A role for the GCC-box in jasmonate-mediated activation of the PDF1.2 gene of *Arabidopsis*. *Plant Physiol.* 132: 1020-1032.
- Brunaud V, Balzergue S, Dubreucq B, Aubourg S, Samson F, Chauvin S, Bechtold N, Cruaud C, DeRose R, Pelletier G, Lepiniec L, Caboche M & Lecharny A** (2002) T-DNA integration into the *Arabidopsis* genome depends on sequences of pre-insertion sites. *EMBO Rep.* 3: 1152-1157.
- Bustin SA** (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol.* 25: 169-193.
- Bustin SA** (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol.* 29: 23-39.
- Callis J, Raasch JA & Vierstra RD** (1990) Ubiquitin extension proteins of *Arabidopsis thaliana*. Structure, localization, and expression of their promoters in transgenic tobacco. *J Biol Chem.* 265: 12486-12493.
- Campbell BR, Song Y, Posch TE, Cullis CA & Town CD** (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene.* 112: 225-228.
- Carpenter JL, Kopczak SD, Snustad DP & Silflow CD** (1993) Semi-constitutive expression of an *Arabidopsis thaliana* alpha-tubulin gene. *Plant Mol Biol.* 21: 937-942.
- Carpenter JL, Ploense SE, Snustad DP & Silflow CD** (1992) Preferential expression of an alpha-tubulin gene of *Arabidopsis* in pollen. *Plant Cell.* 4: 557-571.
- Carteaux F, Thibaud MC, Zimmerli L, Lessard P, Sarrobert C, David P, Gerbaud A, Robaglia C, Somerville S & Nussaume L** (2003) Transcriptome analysis of *Arabidopsis* colonized by a plant-growth promoting rhizobacterium reveals a general effect on disease resistance. *Plant J.* 36: 177-188.
- Castelli V, Aury JM, Jaillon O, Wincker P, Clepet C, Menard M, Cruaud C, Quetier F, Scarpelli C, Schachter V, Temple G, Caboche M, Weissenbach J & Salanoubat M** (2004) Whole genome sequence comparisons and "full-length" cDNA sequences: a combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res.* 14: 406-413.
- CeSC** (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium. *Science.* 282: 2012-2018.
- Chae HS, Faure F & Kieber JJ** (2003) The eto1, eto2, and eto3 mutations and cytokinin treatment increase ethylene biosynthesis in *Arabidopsis* by increasing the stability of ACS protein. *Plant Cell.* 15: 545-559.
- Chamot D, Magee WC, Yu E & Owtrim GW** (1999) A cold shock-induced cyanobacterial RNA helicase. *J Bacteriol.* 181: 1728-1732.
- Charrier B, Champion A, Henry Y & Kreis M** (2002) Expression profiling of the whole *Arabidopsis* shaggy-like kinase multigene family by real-time reverse transcriptase-polymerase chain reaction. *Plant Physiol.* 130: 577-590.
- Chen JJ, Lee S, Zhou G, Rowley JD & Wang SM** (2003) Generation of longer cDNA fragments from SAGE tags for gene identification. *Methods Mol Biol.* 221: 207-222.
- Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Z, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X & Zhu T** (2002) Expression profile matrix

of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell*. 14: 559-574.

**Cheng Z, Snustad DP & Carter JV** (2001) Temporal and spatial expression patterns of *TUB9*, a beta-tubulin gene of *Arabidopsis thaliana*. *Plant Mol Biol*. 47: 389-398.

**Chien A, Edgar DB & Trela JM** (1976) Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J Bacteriol*. 127: 1550-1557.

**Choi DW, Rodriguez EM & Close TJ** (2002) Barley Cbf3 gene identification, expression pattern, and map location. *Plant Physiol*. 129: 1781-1787.

**Chu B, Wilson TJ, McCune-Zierath C, Snustad DP & Carter JV** (1998) Two beta-tubulin genes, *TUB1* and *TUB8*, of *Arabidopsis* exhibit largely nonoverlapping patterns of expression. *Plant Mol Biol*. 37: 785-790.

**Collakova E & DellaPenna D** (2003) The role of homogentisate phytyltransferase and other tocopherol pathway enzymes in the regulation of tocopherol synthesis during abiotic stress. *Plant Physiol*. 133: 930-940.

**Coller JM, Tucker M, Sheth U, Valencia-Sanchez MA & Parker R** (2001) The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA*. 7: 1717-1727.

**Cong B, Liu JP & Tanksley SD** (2002) Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc Natl Acad Sci U S A*. 99: 13606-13611.

**Cooke J, Nowak MA, Boerlijst M & Maynard-Smith J** (1997) Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genet*. 13: 360-364.

**Cooke R, Raynal M, Laudie M, Grellet F, Delseny M, Morris PC, Guerrier D, Giraudat J, Quigley F, Clabault G, Li YF, Mache R, Krivitzky M, Gy IJ, Kreis M, Lecharny A, Parmentier Y, Marbach J, Fleck J, Clement B, Philipps G, Herve C, Bardet C, Tremousaygue D, Höfte H & al.** (1996) Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5000 non-redundant ESTs. *Plant J*. 9: 101-124.

**Cooper B** (2001) Collateral gene expression changes induced by distinct plant viruses during the hypersensitive resistance reaction in *Chenopodium amaranticolor*. *Plant J*. 26: 339-349.

**Courage-Tebbe U, Doring HP, Fedoroff N & Starlinger P** (1983) The controlling element Ds at the Shrunken locus in *Zea mays*: structure of the unstable sh-m5933 allele and several revertants. *Cell*. 34: 383-393.

**Crowe ML, Serizet C, Thareau V, Aubourg B, Rouze P, Hilson P, Beynon J, Weisbeek P, van Hummelen P, Raymond P, Paz Ares J, Niefeld W & Trick M** (2003) CATMA: a complete *Arabidopsis* GST database. *Nucleic Acids Res*. 31: 156-158.

**Craigon DJ, James N, Okyere J, Higgins J, Jotham J & May S** (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res*. 32: D575-577.

**Curie C, Axelos M, Bardet C, Atanassova R, Chaubet N & Lescure B** (1993) Modular organization and development activity of an *Arabidopsis thaliana* *EF-1 alpha* gene promoter. *Mol Gen Genet*. 238: 428-436.

**Curie C, Liboz T, Bardet C, Gander E, Medale C, Axelos M & Lescure B** (1991) Cis and trans-acting elements involved in the activation of *Arabidopsis thaliana* A1 gene encoding the translation elongation factor EF-1 alpha. *Nucleic Acids Res*. 19: 1305-1310.

**Dambrauskas G, Aves SJ, Bryant JA, Francis D & Rogers HJ** (2003) Genes encoding two essential DNA replication activation proteins, Cdc6 and Mcm3, exhibit very different patterns of expression in the tobacco BY-2 cell cycle. *J Exp Bot*. 54: 699-706.

**Datson NA, van der Perk-de Jong J, van den Berg MP, de Kloet ER & Vreugdenhil E** (1999) MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res*. 27: 1300-1307.

**De la Cruz J, Kressler D & Linder P** (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Biochem Sci*. 24: 192-198.

**De Los Reyes BG, Morsy M, Gibbons J, Varma TS, Antoine W, McGrath JM, Halgren R & Redus M** (2003) A snapshot of the low temperature stress transcriptome of developing rice seedlings (*Oryza sativa* L.) via ESTs from subtracted cDNA library. *Theor Appl Genet*. 107: 1071-1082.

**De Valoir T, Tucker MA, Belikoff EJ, Camp LA, Bolduc C & Beckingham K** (1991) A second maternally expressed *Drosophila* gene encodes a putative RNA helicase of the "DEAD box" family. *Proc Natl Acad Sci U S A*. 88: 2113-2117.

**Delseny M** (2003) Towards an accurate sequence of the rice genome. *Curr Opin Plant Biol*. 6: 101-105.

**Demura T, Tashiro G, Horiguchi G, Kishimoto N, Kubo M, Matsuoka N, Minami A, Nagata-Hiwatashi M, Nakamura K, Okamura Y, Sassa N, Suzuki S, Yazaki J, Kikuchi S & Fukuda H** (2002) Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells. *Proc Natl Acad Sci U S A*. 99: 15794-15799.

**Dieci G & Sentenac A** (2003) Detours and shortcuts to transcription reinitiation. *Trends Biochem Sci*. 28: 202-209.

**Donson J, Fang YW, Espiritu Santo G, Xing WM, Salazar A, Miyamoto S, Armendarez V & Volkmuth W** (2002) Comprehensive gene expression analysis by transcript profiling. *Plant Mol Biol*. 48: 75-97.

- Doohan FM, Weston G, Rezanoor HN, Parry DW & Nicholson P** (1999) Development and use of a reverse transcription-PCR assay to study expression of Tri5 by *Fusarium* species in vitro and in planta. *Appl Environ Microbiol.* 65: 3850-3854.
- Du Z, Scott AD & May GD** (2003) Amplification of high-quantity serial analysis of gene expression ditags and improvement of concatemer cloning efficiency. *Biotechniques.* 35: 66-67, 70-62.
- Duque P & Chua NH** (2003) IMB1, a bromodomain protein induced during seed imbibition, regulates ABA- and phyA-mediated responses of germination in *Arabidopsis*. *Plant J.* 35: 787-799.
- Eberl DF, Lorenz LJ, Melnick MB, Sood V, Lasko P & Perrimon N** (1997) A new enhancer of position-effect variegation in *Drosophila melanogaster* encodes a putative RNA helicase that binds chromosomes and is regulated by the cell cycle. *Genetics.* 146: 951-963.
- Edgar R, Domrachev M, Lash AEF-C & Mouse Genome Database G** (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30: 207-210.
- Edwalds-Gilbert G, Veraldi KL & Milcarek C** (1997) Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* 25: 2547-2561.
- Edwards SG, Pirgozliev SR, Hare MC & Jenkinson P** (2001) Quantification of trichothecene-producing *Fusarium* species in harvested grain by competitive PCR to determine efficacies of fungicides against fusarium head blight of winter wheat. *Appl Environ Microbiol.* 67: 1575-1580.
- Ekker M, Wegner J, Akimenko MA & Westerfield M** (1992) Coordinate embryonic expression of three zebrafish *engrailed* genes. *Development.* 116: 1001-1010.
- Ekman DR, Lorenz WW, Przybyla AE, Wolfe NL & Dean JF** (2003) SAGE analysis of transcriptome responses in *Arabidopsis* roots exposed to 2,4,6-trinitrotoluene. *Plant Physiol.* 133: 1397-1406.
- Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA & Liotta LA** (1996) Laser capture microdissection. *Science.* 274: 998-1001.
- Engels WR** (1993) Contributing software to the internet: the Amplify program. *Trends Biochem Sci.* 18: 448-450.
- Estelle MA & Somerville JE** (1986) The mutants of *Arabidopsis*. *Trends Genet.* 2: 89-93.
- Ewing B & Green P** (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat Genet.* 25: 232-234.
- Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S & Claverie JM** (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. *Genome Res.* 9: 950-959.
- Faccioli P, Lagonigro MS, De Cecco L, Stanca AM, Alberici R & Terzi V** (2002) Analysis of differential expression of barley ESTs during cold acclimatization using microarray technology. *Plant Biol.* 4: 630-639.
- Featherstone M** (2002) Coactivators in transcription initiation: here are your orders. *Curr Opin Genet Dev.* 12: 149-155.
- Feldbrugge M, Sprenger M, Hahlbrock K & Weisshaar B** (1997) PcMYB1, a novel plant protein containing a DNA-binding domain with one MYB repeat, interacts in vivo with a light-regulatory promoter unit. *Plant J.* 11: 1079-1093.
- Ferea TL, Botstein D, Brown PO & Rosenzweig RF** (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci U S A.* 96: 9721-9726.
- Ferre F** (1992) Quantitative or semi-quantitative PCR: reality versus myth. *PCR Methods Appl.* 2: 1-9.
- Ferris SD & Whitt GS** (1979) Evolution of the differential regulation of duplicate genes after polyploidization. *J Mol Evol.* 12: 267-317.
- Filipowicz W, Pelczar P, Pogacic V & Dragon F** (1999) Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. *Acta Biochim Pol.* 46: 377-389.
- Fischer N & Weis K** (2002) The DEAD box protein Dhh1 stimulates the decapping enzyme Dcp1. *EMBO J.* 21: 2788-2797.
- Fizames C, Munos S, Cazettes C, Nacry P, Boucherez J, Gaymard F, Piquemal D, Delorme V, Commes T, Doumas P, Cooke R, Marti J, Sentenac H & Gojon A** (2004) The *Arabidopsis* root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol.* 134: 67-80.
- Florea L, Hartzell G, Zhang Z, Rubin GM & Miller W** (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* 8: 967-974.
- Folta KM, Pontin MA, Karlin Neumann G, Bottini R & Spalding EP** (2003) Genomic and physiological studies of early cryptochrome 1 action demonstrate roles for auxin and gibberellin in the control of hypocotyl growth by blue light. *Plant J.* 36: 203-214.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL & Postlethwait J** (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 151: 1531-1545.
- Foucher F, Morin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ & Rameau C** (2003) DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. *Plant Cell.* 15: 2742-2754.

- Frary A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu JP, Meller J, Elber R, Alpert KB & Tanksley SD** (2000) *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science*. 289: 85-88.
- Freeman WM, Walker SJ & Vrana KE** (1999) Quantitative RT-PCR: pitfalls and potential. *Biotechniques*. 26: 112-122, 124-115.
- Fu H, Kim SY & Park WD** (1995a) High-level tuber expression and sucrose inducibility of a potato *Sus4* sucrose synthase gene require 5' and 3' flanking sequences and the leader intron. *Plant Cell*. 7: 1387-1394.
- Fu H, Kim SY & Park WD** (1995b) A potato *Sus3* sucrose synthase gene contains a context-dependent 3' element and a leader intron with both positive and negative tissue-specific effects. *Plant Cell*. 7: 1395-1403.
- Fujimura M & Takamura K** (2000) Characterization of an ascidian DEAD-box gene, Ci-DEAD1: specific expression in the germ cells and its mRNA localization in the posterior-most blastomeres in early embryos. *Dev Genes Evol*. 210: 64-72.
- Fujiwara Y, Komiya T, Kawabata H, Sato M, Fujimoto H, Furusawa M & Noce T** (1994) Isolation of a DEAD-family protein gene that encodes a murine homolog of *Drosophila* vasa and its specific expression in germ cell lineage. *Proc Natl Acad Sci U S A*. 91: 12258-12262.
- Gachon C, Mingam A & Charrier B** (2004) Real-time PCR: What relevance to plant studies? *J Exp Bot*. soumis
- Gausung K & Barkardottir R** (1986) Structure and expression of ubiquitin genes in higher plants. *Eur J Biochem*. 158: 57-62.
- Gerstein M & Jansen R** (2000) The current excitement in bioinformatics-analysis of whole-genome expression data: how does it relate to protein structure and function? *Curr Opin Struct Biol*. 10: 574-584.
- Gherbi R & Herisson J** (2002) Representation and processing of complex DNA spatial architecture and its annotated genomic content. *Pac Symp Biocomput*. 151-162.
- Gibbings JG, Cook BP, Dufault MR, Madden SL, Khuri S, Turnbull CJ & Dunwell JM** (2003) Global transcript analysis of rice leaf and seed using SAGE technology. *Plant Biotechnol J*. 1: 271-285.
- Gilliland G, Perrin S, Blanchard K & Bunn HF** (1990) Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc Natl Acad Sci U S A*. 87: 2725-2729.
- Girke T, Todd J, Ruuska S, White J, Benning C & Ohlrogge J** (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol*. 124: 1570-1581.
- Goda H, Shimada Y, Asami T, Fujioka S & Yoshida S** (2002) Microarray analysis of brassinosteroid-regulated genes in *Arabidopsis*. *Plant Physiol*. 130: 1319-1334.
- Golem S & Culver JN** (2003) Tobacco mosaic virus induced alterations in the gene expression profile of *Arabidopsis thaliana*. *Mol Plant Microbe Interact*. 16: 681-688.
- Gomez C** (1999) *Engineering and Scientific Computing with Scilab*, Birkhauser, Boston.
- Gong Z, Lee H, Xiong L, Jagendorf A, Stevenson B & Zhu JK** (2002) RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. *Proc Natl Acad Sci U S A*. 99: 11507-11512.
- Goodman HM, Ecker JR & Dean C** (1995) The genome of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 92: 10831-10835.
- Gorbalenya AE, Koonin EV, Donchenko AP & Blinov VM** (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res*. 17: 4713-4730.
- Gu Z, Nicolae D, Lu HH & Li WH** (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet*. 18: 609-613.
- Guescini M, Pierleoni R, Palma F, Zeppa S, Vallorani L, Potenza L, Sacconi C, Giomaro G & Stocchi V** (2003) Characterization of the *Tuber borchii* nitrate reductase gene and its role in ectomycorrhizae. *Mol Genet Genomics*. 1: 1.
- Gutierrez RA, Ewing RM, Cherry JM & Green PJ** (2002) Identification of unstable transcripts in *Arabidopsis* by cDNA microarray analysis: Rapid decay is associated with a group of touch- and specific clock-controlled genes. *Proc Nat Acad Sci U S A*. 99: 11513-11518.
- Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O & Salzberg SL** (2002) Full-length messenger RNA sequences greatly improve genome annotation. *Genome Biol*. 3: research00291-research002912.
- Halterman DA, Wei FS & Wise RP** (2003) Powdery mildew-induced Mla mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiol*. 131: 558-567.
- Hammond JP, Bennett MJ, Bowen HC, Broadley MR, Eastwood DC, May ST, Rahn C, Swarup R, Woolaway KE & White PJ** (2003) Changes in gene expression in *Arabidopsis* shoots during phosphate starvation and the potential for developing smart plants. *Plant Physiol*. 132: 578-596.

- Hanano S, Amagai M, Kaneko T, Kuwata C, Tsugane T, Sakurai N, Nakamura Y, Shibata D & Tabata S** (2002) Analysis of gene expression in *Arabidopsis thaliana* by array hybridization with genomic DNA fragments aligned along chromosomal regions. *Plant J.* 30: 247-255.
- Harari-Steinberg O, Ohad I & Chamovitz DA** (2001) Dissection of the light signal transduction pathways regulating the two early light-induced protein genes in *Arabidopsis*. *Plant Physiol.* 127: 986-997.
- Harmer SL, Hogenesch LB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA & Kay SA** (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science.* 290: 2110-2113.
- Harper JF, Manney L & Sussman MR** (1994) The plasma membrane H(+)-ATPase gene family in *Arabidopsis*: genomic sequence of AHA10 which is expressed primarily in developing seeds. *Mol Gen Genet.* 244: 572-587.
- Heid CA, Stevens J, Livak KJ & Williams PM** (1996) Real time quantitative PCR. *Genome Res.* 6: 986-994.
- Henikoff S & Henikoff JG** (1991) Automated assembly of protein blocks for database searching. *Nucleic Acids Res.* 19: 6565-6572.
- Henry Y, Champion A, Gy I, Picaud A, Lecharny A & Kreis M** (2002) A model for the evolution and genesis of the pseudotetraploid *Arabidopsis thaliana* genome. *Adv Bot Res.* 38: 235-249.
- Hernandez M, Pla M, Esteve T, Prat S, Puigdomenech P & Ferrando A** (2003) A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard based on the 3'-transgene integration sequence. *Transgenic Res.* 12: 179-189.
- Hernandez M, Rio A, Esteve T, Prat S & Pla M** (2001) A rapeseed-specific gene, Acetyl-CoA carboxylase, can be used as a reference for qualitative and real-time quantitative PCR detection of transgenes from mixed food samples. *J Agric Food Chem.* 49: 3622-3627.
- Heuser T & Zimmer W** (2002) Quantitative analysis of phytopathogenic ascomycota on leaves of pedunculate oaks (*Quercus robur* L.) by real-time PCR. *FEMS Microbiol Lett.* 209: 295-299.
- Higuchi R, Fockler C, Dollinger G & Watson R** (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y).* 11: 1026-1030.
- Hillier LD, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, Dietrich N, DuBuque T, Favello A, Gish W, Hawkins M, Hultman M, Kucaba T, Lacy M, Le M, Le N, Mardis E, Moore B, Morris M, Parsons J, Prange C, Rifkin L, Rohlfing T, Schellenberg K, Marra M & et al.** (1996) Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* 6: 807-828.
- Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao R, Bitton F, Caboche M, Cannoot B, Chardakov V, Cognet-Holliger C, Colot V, Crowe M, Darimont C, Durinck S, Eickhoff H, Falcon de Longuevalle A, Farmer EE, Grant M, Kuiper MTR, Lehrach H, Léon C, Leyva A, Lundeberg J, Lurin C, Moreau Y, Nietfeld W, Paz-Ares J, Reymond P, Rouzé P, Sandberg G, Segura MD, Serizet C, Tabrett A, Taconnat L, Thareau V, Van Hummelen P, Vercruyse S, Vuylsteke M, Weingartner M, Weisbeek PJ, Wirta V, Wittink FRA, Zabeau M & Small I** (2004) Versatile Gene-Specific Sequence Tags for *Arabidopsis* Functional Genomics: transcript profiling and reverse genetics applications. *Genome Res.* In press.
- Höfte H, Desprez T, Amselem J, Chiapello H, Rouze P, Caboche M, Moisan A, Jourjon MF, Charpentreau JL, Berthomieu P & et al.** (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* 4: 1051-1061.
- Honys D & Twell D** (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. *Plant Physiol.* 132: 640-652.
- Horvath DP, Schaffer R, West M & Wisman E** (2003a) *Arabidopsis* microarrays identify conserved and differentially expressed genes involved in shoot growth and development from distantly related plant species. *Plant J.* 34: 125-134.
- Horvath DP, Schaffer R & Wisman E** (2003b) Identification of genes induced in emerging tillers of wild oat (*Avena fatua*) using *Arabidopsis* microarrays. *Weed Sci.* 51: 503-508.
- Hoth S, Morgante M, Sanchez JP, Hanafey MK, Tingey SV & Chua NH** (2002) Genome-wide gene expression profiling in *Arabidopsis thaliana* reveals new targets of abscisic acid and largely impaired gene regulation in the *abi1-1* mutant. *J Cell Sci.* 115: 4891-4900.
- Hough AJ, Harbison SA, Savill MG, Melton LD & Fletcher G** (2002) Rapid enumeration of *Listeria monocytogenes* in artificially contaminated cabbage using real-time polymerase chain reaction. *J Food Prot.* 65: 1329-1332.
- Houlne G & Boutry M** (1994) Identification of an *Arabidopsis thaliana* gene encoding a plasma membrane H(+)-ATPase whose expression is restricted to anther tissue. *Plant J.* 5: 311-317.
- Huang S, An YQ, McDowell JM, McKinney EC & Meagher RB** (1996) The *Arabidopsis thaliana* ACT4/ACT12 actin gene subclass is strongly expressed throughout pollen development. *Plant J.* 10: 189-202.
- Huang S, An YQ, McDowell JM, McKinney EC & Meagher RB** (1997) The *Arabidopsis* ACT11 actin gene is strongly expressed in tissues of the emerging inflorescence, pollen, and developing ovules. *Plant Mol Biol.* 33: 125-139.



- Hubner P, Waiblinger HU, Pietsch K & Brodmann P** (2001) Validation of PCR methods for quantitation of genetically modified plants in food. *J AOAC Int.* 84: 1855-1864.
- Hughes MK & Hughes AL** (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol Biol Evol.* 10: 1360-1369.
- Ilan J** (1977) Unwinding protein specific for mRNA translation fractionated together with rabbit reticulocyte initiation factor 3 complex. *Proc Natl Acad Sci U S A.* 74: 2325-2329.
- Ingham DJ, Beer S, Money S & Hansen G** (2001) Quantitative real-time PCR assay for determining transgene copy number in transformed plants. *Biotechniques.* 31: 132-134, 136-140.
- Itadani H, Sugita M & Sugiura M** (1994) Structure and expression of a cDNA encoding an RNA helicase-like protein in tobacco. *Plant Mol Biol.* 24: 249-252.
- Ito T, Sakai H & Meyerowitz EM** (2003) Whorl-specific expression of the SUPERMAN gene of *Arabidopsis* is mediated by cis elements in the transcribed region. *Curr Biol.* 13: 1524-1530.
- Itoh R, Fujiwara M, Nagata N & Yoshida S** (2001) A chloroplast protein homologous to the eubacterial topological specificity factor MinE plays a role in chloroplast division. *Plant Physiol.* 127: 1644-1655.
- Jacobs AK, Lipka V, Burton RA, Panstruga R, Strizhov N, Schulze-Lefert P & Fincher GB** (2003) An *Arabidopsis* callose synthase, GSL5, is required for wound and papillary callose formation. *Plant Cell.* 15: 2503-2513.
- Jaramillo M, Browning K, Dever TE, Blum S, Trachsel H, Merrick WC, Ravel JM & Sonenberg N** (1990) Translation initiation factors that function as RNA helicases from mammals, plants and yeast. *Biochim Biophys Acta.* 1050: 134-139.
- Jensen TH, Boulay J, Rosbash M & Libri D** (2001) The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr Biol.* 11: 1711-1715.
- Jones PG, Allaway D, Gilmour DM, Harris C, Rankin D, Retzel ER & Jones CA** (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. *Planta.* 216: 255-264.
- Joshi CP** (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res.* 15: 6643-6653.
- Jung SH, Lee JY & Lee DH** (2003) Use of SAGE technology to reveal changes in gene expression in *Arabidopsis* leaves undergoing cold stress. *Plant Mol Biol.* 52: 553-567.
- Kandasamy MK, Gilliland LU, McKinney EC & Meagher RB** (2001) One plant actin isoform, ACT7, is induced by auxin and required for normal callus formation. *Plant Cell.* 13: 1541-1554.
- Kellogg EA** (2001) Root hairs, trichomes and the evolution of duplicate genes. *Trends Plant Sci.* 6: 550-552.
- Kenzelmann M & Muhlemann K** (1999) Substantially enhanced cloning efficiency of SAGE (Serial Analysis of Gene Expression) by adding a heating step to the original protocol. *Nucleic Acids Res.* 27: 917-918.
- Kerk NM, Ceserani T, Tausta SL, Sussex IM & Nelson TM** (2003) Laser capture microdissection of cells from plant tissues. *Plant Physiol.* 132: 27-35.
- Kim Y & An G** (1992) Pollen-specific expression of the *Arabidopsis thaliana* alpha 1-tubulin promoter assayed by beta-glucuronidase, chloramphenicol acetyltransferase and diphtheria toxin reporter genes. *Transgenic Res.* 1: 188-194.
- Klahre U, Friederich E, Kost B, Louvard D & Chua NH** (2000) Villin-like actin-binding proteins are expressed ubiquitously in *Arabidopsis*. *Plant Physiol.* 122: 35-48.
- Klein D** (2002) Quantification using real-time PCR technology: applications and limitations. *Trends Mol Med.* 8: 257-260.
- Klok EJ, Wilson IW, Wilson D, Chapman SC, Ewing RM, Somerville SC, Peacock WJ, Dolferus R & Dennis ES** (2002) Expression profile analysis of the low-oxygen response in *Arabidopsis* root cultures. *Plant Cell.* 14: 2481-2494.
- Koo AJK & Ohlrogge JB** (2002) The predicted candidates of *Arabidopsis* plastid inner envelope membrane proteins and their expression profiles. *Plant Physiol.* 130: 823-836.
- Kopczak SD, Haas NA, Hussey PJ, Silflow CD & Snustad DP** (1992) The small genome of *Arabidopsis* contains at least six expressed alpha-tubulin genes. *Plant Cell.* 4: 539-547.
- Kosugi S & Ohashi Y** (1997) PCF1 and PCF2 specifically bind to cis-elements in the rice proliferating cell nuclear antigen gene. *Plant Cell.* 9: 1607-1619.
- Kosugi S & Ohashi Y** (2002) DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant J.* 30: 337-348.
- Kosugi S, Suzuka I & Ohashi Y** (1995) Two of three promoter elements identified in a rice gene for proliferating cell nuclear antigen are essential for meristematic tissue-specific expression. *Plant J.* 7: 877-886.
- Kotake T, Takada S, Nakahigashi K, Ohto M & Goto K** (2003) *Arabidopsis* TERMINAL FLOWER 2 gene encodes a heterochromatin protein 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes. *Plant Cell Physiol.* 44: 555-564.
- Kowalski SP, Lan TH, Feldmann KA & Paterson AH** (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics.* 138: 499-510.

- Kressler D, de la Cruz J, Rojo M & Linder P** (1997) Fallp is an essential DEAD-box protein involved in 40S-ribosomal- subunit biogenesis in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 17: 7283-7294.
- Krizek BA, Prost V, Joshi RM, Stoming T & Glenn TC** (2003) Developing transgenic *Arabidopsis* plants to be metal-specific bioindicators. *Environ Toxicol Chem*. 22: 175-181.
- Ku HM, Vision T, Liu JP & Tanksley SD** (2000) Comparing sequenced segments of the tomato and *Arabidopsis* genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc Natl Acad Sci U S A*. 97: 9121-9126.
- Kuribara H, Shindo Y, Matsuoka T, Takubo K, Futo S, Aoki N, Hirao T, Akiyama H, Goda Y, Toyoda M & Hino A** (2002) Novel reference molecules for quantitation of genetically modified maize and soybean. *J AOAC Int*. 85: 1077-1089.
- Kursteiner O, Dupuis I & Kuhlemeier C** (2003) The pyruvate decarboxylase1 gene of *Arabidopsis* is required during anoxia but not other environmental stresses. *Plant Physiol*. 132: 968-978.
- Kurth J, Varotto C, Pesaresi P, Biehl A, Richly E, Salamini F & Leister D** (2002) Gene-sequence-tag expression analyses of 1,800 genes related to chloroplast functions. *Planta*. 215: 101-109.
- Lake V & Willows RD** (2003) Rapid extraction of RNA and analysis of transcript levels in *Chlamydomonas reinhardtii* using real-time RT-PCR: Magnesium chelatase chlH, chlD and chlI gene expression. *Photosynth Res*. 77: 69-76.
- Lan TH, DelMonte TA, Reischmann KP, Hyman J, Kowalski SP, McFerson J, Kresovich S & Paterson AH** (2000) An EST-enriched comparative map of *Brassica oleracea* and *Arabidopsis thaliana*. *Genome Res*. 10: 776-788.
- Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO & Davis RW** (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A*. 94: 13057-13062.
- Laule O, Furlholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Grussem W & Lange BM** (2003) Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc Nat Acad Sci U S A*. 100: 6866-6871.
- Le Guen L, Thomas M & Kreis M** (1994) Gene density and organization in a small region of the *Arabidopsis thaliana* genome. *Mol Gen Genet*. 245: 390-396.
- Lecharyn A, Boudet N, Gy I, Aubourg S & Kreis M** (2003) Introns in, introns out in plant genes families: a genomic approach of the dynamics of gene structure. *J Struct Funct Genomics*. 3: 111-116.
- Lechelt-Kunze C, Meissner RC, Drewes M & Tietjen K** (2003) Flufenacet herbicide treatment phenocopies the fiddlehead mutant in *Arabidopsis thaliana*. *Pest Manag Sci*. 59: 847-856.
- Lee JY & Lee DH** (2003) Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. *Plant Physiol*. 132
- Lee S, Clark T, Chen J, Zhou G, Scott LR, Rowley JD & Wang SM** (2002) Correct identification of genes from serial analysis of gene expression tag sequences. *Genomics*. 79: 598-602.
- Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P & Rombauts S** (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 30: 325-327.
- Li SC, Chung MC & Chen CS** (2001a) Cloning and characterization of a DEAD box RNA helicase from the viable seedlings of aged mung bean. *Plant Mol Biol*. 47: 761-770.
- Li W, Belsham GJ & Proud CG** (2001b) Eukaryotic initiation factors 4A (eIF4A) and 4G (eIF4G) mutually interact in a 1 : 1 ratio in vivo. *J Biol Chem*. 276: 29111-29115.
- Liboz T, Bardet C, Le Van Thai A, Axelos M & Lescure B** (1990) The four members of the gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1 alpha are actively transcribed. *Plant Mol Biol*. 14: 107-110.
- Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, Nishi K, Schnier J & Slonimski PP** (1989) Birth of the D-E-A-D box. *Nature*. 337: 121-122.
- Linder P & Slonimski PP** (1989) An essential yeast protein, encoded by duplicated genes TIF1 and TIF2 and homologous to the mammalian translation initiation factor eIF-4A, can suppress a mitochondrial missense mutation. *Proc Natl Acad Sci U S A*. 86: 2286-2290.
- Linder P & Stutz F** (2001) mRNA export: Travelling with DEAD box proteins. *Curr Biol*. 11: R961-R963.
- Linder P, Tanner NK & Banroques J** (2001) From RNA helicases to RNPases. *Trends Biochem Sci*. 26: 339-341.
- Liu JP, Cong B & Tanksley SDC-R-A-B** (2003) Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the cloned quantitative trait locus *fw2.2* controls fruit size. *Plant Physiol*. 132: 292-299.
- Locatelli G, Urso V & Malnati M** (2000) Quantitative analysis of GMO food contaminations using real time PCR. *Ital J Biochem*. 49: 61-63.

- Lombardi MP, van den Hoff MJ, Ruijter JM, Luijterink M, Buffing AA, Markman MW, Moorman AF & Lekanne Deprez RH** (2003) Expression analysis of subtractively enriched libraries (EASEL): a widely applicable approach to the identification of differentially expressed genes. *J Biochem Biophys Methods*. 57: 17-33.
- Long EO & Dawid IB** (1980) Repeated genes in eukaryotes. *Annu Rev Biochem*. 49: 727-764.
- Lorenz WW & Dean JF** (2002) SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). *Tree Physiol*. 22: 301-310.
- Lorkovic ZJ, Herrmann RG & Oelmüller R** (1997) PRH75, a new nucleus-localized member of the DEAD-box protein family from higher plants. *Mol Cell Biol*. 17: 2257-2265.
- Ludwig SR, Oppenheimer DG, Silflow CD & Snustad DP** (1987) Characterization of the alpha-tubulin gene family of *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 84: 5833-5837.
- Lüking A, Stahl U & Schmidt U** (1998) The protein family of RNA helicases. *Crit Rev Biochem Mol Biol*. 33: 259-296.
- Lynch M & Conery JS** (2000) The evolutionary fate and consequences of duplicate genes. *Science*. 290: 1151-1155.
- Lynch M & Force A** (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics*. 154: 459-473.
- Lynch M, O'Hely M, Walsh B & Force A** (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics*. 159: 1789-1804.
- Maathuis FJM, Filatov V, Herzyk P, Krijger GC, Axelsen KB, Chen SX, Green BJ, Li Y, Madagan KL, Sanchez Fernandez R, Forde BG, Palmgren MG, Rea PA, Williams LE, Sanders D & Amtmann A** (2003) Transcriptome analysis of root transporters reveals participation of multiple gene families in the response to cation stress. *Plant J*. 35: 675-692.
- MacArthur H, Houston DW, Bubunenko M, Mosquera L & King ML** (2000) DEADSOUTH is a germ plasm specific DEAD-box RNA helicase in *Xenopus* related to eIF4A. *Mech Dev*. 95: 291-295.
- Maden BE & Hughes JM** (1997) Eukaryotic ribosomal RNA: the recent excitement in the nucleotide modification problem. *Chromosoma*. 105: 391-400.
- Maguire TL, Grimmond S, Forrest A, Iturbe Ormaetxe I, Meksem K & Gresshoff P** (2002) Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis. *J Plant Physiol*. 159: 1361-1374.
- Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL & Dietrich RA** (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nat Genet*. 26: 403-410.
- Mandel T, Fleming AJ, Krahenbuhl R & Kuhlemeier C** (1995) Definition of constitutive gene expression in plants: the translation initiation factor 4A gene as a model. *Plant Mol Biol*. 29: 995-1004.
- Manevski A, Bardet C, Tremousaygue D & Lescure B** (1999) Characterization and properties of heteromeric plant protein complexes that interact with *tef cis*-acting elements in both RNA polymerase II- dependent promoters and rDNA spacer sequences. *Mol Gen Genet*. 261: 892-900.
- Manevski A, Bertoni G, Bardet C, Tremousaygue D & Lescure B** (2000) In synergy with various *cis*-acting elements, plant interstitial telomere motifs regulate gene expression in *Arabidopsis* root meristems. *FEBS Lett*. 483: 43-46.
- Manley JL** (2002) Nuclear coupling: RNA processing reaches back to transcription. *Nat Struct Biol*. 9: 790-791.
- Marker C, Zemann A, Terhorst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie JP, Brosius J & Huttenhofer A** (2002a) Experimental RNomics. Identification of 140 Candidates for Small Non- Messenger RNAs in the Plant *Arabidopsis thaliana*. *Curr Biol*. 12: 2002-2013.
- Marker C, Zemann A, Terhorst T, Kiefmann M, Kastenmayer JP, Green P, Bachellerie JP, Brosius J & Huttenhofer A** (2002b) Experimental RNomics. Identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr Biol*. 12: 2002-2013.
- Martinez IM & Chrispeels MJ** (2003) Genomic analysis of the unfolded protein response in *Arabidopsis* shows its connection to important cellular processes. *Plant Cell*. 15: 561-576.
- Mason G, Provero P, Vaira AM & Accotto GP** (2002) Estimating the number of integrations in transformed plants by quantitative real-time PCR. *BMC Biotechnol*. 2: 20.
- Matsumura H, Nirasawa S, Kiba A, Urasaki N, Saitoh H, Ito M, Kawai-Yamada M, Uchimiya H & Terauchi R** (2003) Overexpression of Bax inhibitor suppresses the fungal elicitor-induced cell death in rice (*Oryza sativa* L.) cells. *Plant J*. 33: 425-434.
- Matsumura H, Nirasawa S & Terauchi R** (1999) Transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). *Plant J*. 20
- McDowell JM, An YQ, Huang S, McKinney EC & Meagher RB** (1996a) The *Arabidopsis ACT7* actin gene is expressed in rapidly developing tissues and responds to several external stimuli. *Plant Physiol*. 111: 699-711.
- McDowell JM, Huang S, McKinney EC, An YQ & Meagher RB** (1996b) Structure and evolution of the actin gene family in *Arabidopsis thaliana*. *Genetics*. 142: 587-602.

- McGinnis KM, Thomas SG, Soule JD, Strader LC, Zale JM, Sun TP & Steber CM** (2003) The *Arabidopsis* SLEEPY1 gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell*. 15: 1120-1130.
- McGrath JM, Jancso MM & Pichersky E** (1993) Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana*. *Theor Appl Genet*. 86: 880-888.
- McMaugh SJ & Lyon BR** (2003) Real-time quantitative RT-PCR assay of gene expression in plant roots during fungal pathogenesis. *Biotechniques*. 34: 982-986.
- Meagher RB, McKinney EC & Vitale AV** (1999) The evolution of new structures: clues from plant cytoskeletal genes. *Trends Genet*. 15: 278-284.
- Meinke DW, Cherry JM, Dean C, Rounsley SD & Koornneef M** (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science*. 282: 662, 679-682.
- Meister RJ, Kotow LM & Gasser CS** (2002) SUPERMAN attenuates positive INNER NO OUTER autoregulation to maintain polar development of *Arabidopsis* ovule outer integuments. *Development*. 129: 4281-4289.
- Meli M, Albert-Fournier B & Maurel MC** (2001) Recent findings in the modern RNA world. *Int Microbiol*. 4: 5-11.
- Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF & Schmidt RJ** (1996) Diversification of C-function activity in maize flower development. *Science*. 274: 1537-1540.
- Menges M, Hennig L, Gruissem W & Murray JA** (2002) Cell cycle-regulated gene expression in *Arabidopsis*. *J Biol Chem*. 277: 41987-42002.
- Metz AM, Timmer RT & Browning KS** (1992) Sequences for two cDNAs encoding *Arabidopsis thaliana* eukaryotic protein synthesis initiation factor 4A. *Gene*. 120: 313-314.
- Meyerowitz EM** (1992) *Introduction to the Arabidopsis genome*, Singapore: World Scientific Publishing Co. Pte. Ltd.
- Meyers BC, Dickerman AW, Micheltore RW, Sivaramakrishnan S, Sobral BW & Young ND** (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J*. 20: 317-332.
- Mingam A, Gachon C & Charrier B** (2003) Quantifying nucleic acids with fluorescent probes: Contribution of real-time PCR to applied and fundamental plant research. *Adv Plant Physiol*. 7: 423-458.
- Mingam A, Toffano-Nioche C, Brunaud V, Boudet N, Kreis M & Lecharny L** (2004) DEAD-box RNA helicases in *Arabidopsis thaliana*: establishing a link between quantitative expression, gene structure and evolution of a family of genes. *Plant Biotechnol J*: accepté.
- Mladek C, Guger K & Hauser MT** (2003) Identification and characterization of the ARIADNE gene family in *Arabidopsis*. A group of putative E3 ligases. *Plant Physiol*. 131: 27-40.
- Moffat AS** (2000) Genetics. Transposons help sculpt a dynamic genome. *Science*. 289: 1455-1457.
- Monroe-Augustus M, Zolman BK & Bartel B** (2003) IBR5, a dual-specificity phosphatase-like protein modulating auxin and abscisic acid responsiveness in *Arabidopsis*. *Plant Cell*. 15: 2979-2991.
- Montrichard F, Renard M, Alkhalfioui F, Duval FD & Macherel D** (2003) Identification and differential expression of two thioredoxin h isoforms in germinating seeds from pea. *Plant Physiol*. 132: 1707-1715.
- Morissette DC, Driscoll BT & Jabaji-Hare S** (2003) Molecular cloning, characterization, and expression of a cDNA encoding an endochitinase gene from the mycoparasite *Stachybotrys elegans*. *Fungal Genet Biol*. 39: 276-285.
- Moseyko N, Zhu T, Chang HS, Wang X & Feldman LJ** (2002) Transcription profiling of the early gravitropic response in *Arabidopsis* using high-density oligonucleotide probe microarrays. *Plant Physiol*. 130: 720-728.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G & Erlich H** (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*. 51: 263-273.
- Mullis KB & Faloona FA** (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*. 155: 335-350.
- Mussig C, Fischer S & Altmann T** (2002) Brassinosteroid-regulated gene expression. *Plant Physiol*. 129: 1241-1251.
- Mussig C, Shin GH & Altmann T** (2003) Brassinosteroids promote root growth in *Arabidopsis*. *Plant Physiol*. 133: 1261-1271.
- Nadeau JH & Sankoff D** (1997) Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. *Genetics*. 147: 1259-1266.
- Nakazono M, Qiu F, Borsuk LA & Schnable PS** (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: Identification of genes expressed differentially in epidermal cells or vascular tissues of maize. *Plant Cell*. 15: 583-596.
- Narusaka Y, Narusaka M, Seki M, Ishida J, Nakashima M, Kamiya A, Enju A, Sakurai T, Satoh M, Kobayashi M, Tosa Y, Park P & Shinozaki K** (2003) The cDNA microarray analysis using an *Arabidopsis* pad3 mutant reveals the expression profiles and classification of genes induced by *Alternaria brassicicola* attack. *Plant Cell Physiol*. 44: 377-387.

- Nasmyth K, Dirick L, Surana U, Amon A & Cvrckova F** (1991) Some facts and thoughts on cell cycle control in yeast. *Cold Spring Harb Symp Quant Biol.* 56: 9-20.
- Navarro RE, Shim EY, Kohara Y, Singson A & Blackwell TK** (2001) cgh-1, a conserved predicted RNA helicase required for gametogenesis and protection from physiological germline apoptosis in *C. elegans*. *Development.* 128: 3221-3232.
- Negishi T, Nakanishi H, Yazaki J, Kishimoto N, Fujii F, Shimbo K, Yamamoto K, Sakata K, Sasaki T, Kikuchi S, Mori S & Nishizawa NK** (2002) cDNA microarray analysis of gene expression during Fe-deficiency stress in barley suggests that polar transport of vesicles is implicated in phytosiderophore secretion in Fe-deficient barley roots. *Plant J.* 30: 83-94.
- Nesbitt TC & Tanksley SD** (2002) Comparative sequencing in the genus *Lycopersicon*: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics.* 162: 365-379.
- Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M & et al.** (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* 106: 1241-1255.
- Niessen ML & Vogel RF** (1998) Group specific PCR-detection of potential trichothecene-producing *Fusarium*-species in pure cultures and cereal samples. *Syst Appl Microbiol.* 21: 618-631.
- Nissan TA, Bassler J, Petfalski E, Tollervey D & Hurt E** (2002) 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. *EMBO J.* 21: 5539-5547.
- Noh SJ, Kwon CS, Oh DH, Moon JS & Chung WI** (2003) Expression of an evolutionarily distinct novel BiP gene during the unfolded protein response in *Arabidopsis thaliana*. *Gene.* 311: 81-91.
- Norris SR, Meyer SE & Callis J** (1993) The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. *Plant Mol Biol.* 21: 895-906.
- Ogawa M, Hanada A, Yamauchi Y, Kuwalhara A, Kamiya Y & Yamaguchi S** (2003) Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *Plant Cell.* 15: 1591-1604.
- Ohno S** (1970) *Evolution by gene duplication*, Spring Verlag New York
- Okanami M, Meshi T & Iwabuchi M** (1998) Characterization of a DEAD box ATPase/RNA helicase protein of *Arabidopsis thaliana*. *Nucleic Acids Res.* 26: 2638-2643.
- Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K, Iida K, Maruyama K, Satoh S, Yamaguchi-Shinozaki K & Shinozaki K** (2003) Monitoring expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. *Plant J.* 34: 868-887.
- op den Camp RGL, Przybyla D, Ochsenbein C, Laloi C, Kim CH, A DA, Wagner D, Hideg E, Gobel C, Feussner I, Nater M & Apel K** (2003) Rapid induction of distinct stress responses after the release of singlet oxygen in *Arabidopsis*. *Plant Cell.* 15: 2320-2332.
- Oppenheimer DG, Haas N, Silflow CD & Snustad DP** (1988) The beta-tubulin gene family of *Arabidopsis thaliana*: preferential accumulation of the beta 1 transcript in roots. *Gene.* 63: 87-102.
- Orsel M, Krapp A & Daniel Vedele F** (2002a) Analysis of the NRT2 nitrate transporter family in *Arabidopsis*. Structure and gene expression. *Plant Physiol.* 129: 886-896.
- Orsel M, Krapp A & Daniel-Vedele F** (2002b) Analysis of the NRT2 nitrate transporter family in *Arabidopsis*. Structure and gene expression. *Plant Physiol.* 129: 886-896.
- Panchuk, II, Volkov RA & Schoffl F** (2002) Heat stress- and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in *Arabidopsis*. *Plant Physiol.* 129: 838-853.
- Papp B, Pal C & Hurst LD** (2003) Evolution of cis-regulatory elements in duplicated genes of yeast. *Trends Genet.* 19: 417-422.
- Paquette SM, Bak S & Feyereisen R** (2000) Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA Cell Biol.* 19: 307-317.
- Park SC, Kwon HB & Shih MC** (1996) Cis-acting elements essential for light regulation of the nuclear gene encoding the A subunit of chloroplast glyceraldehyde 3-phosphate dehydrogenase in *Arabidopsis thaliana*. *Plant Physiol.* 112: 1563-1571.
- Park WJ, Kriechbaumer V, Muller A, Piotrowski M, Meeley RB, Gierl A & Glawischnig E** (2003) The nitrilase ZmNIT2 converts indole-3-acetonitrile to indole-3-acetic acid. *Plant Physiol.* 4: 4.
- Pause A, Methot N & Sonenberg N** (1993) The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol Cell Biol.* 13: 6789-6798.
- Pause A & Sonenberg N** (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* 11: 2643-2654.
- Peeters NM & Hanson MR** (2002) Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA.* 8: 497-511.
- Perez Amador MA, Lidder P, Johnson MA, Landgraf J, Wisman E & Green PJ** (2001) New molecular phenotypes in the *dst* mutants of *Arabidopsis* revealed by DNA microarray analysis. *Plant Cell.* 13: 2703-2717.

- Pfaffl MW** (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29: e45.
- Philippar K, Buchsenschutz K, Abshagen M, Fuchs I, Geiger D, Lacombe B & Hedrich R** (2003) The K<sup>+</sup> channel KZM1 mediates potassium uptake into the phloem and guard cells of the C-4 grass *Zea mays*. *J Biol Chem.* 278: 16973-16981.
- Piatigorsky J** (2003) Crystallin genes: specialization by changes in gene regulation may precede gene duplication. *J Struct Funct Genomics.* 3: 131-137.
- Plesse B, Criqui MC, Durr A, Parmentier Y, Fleck J & Genschik P** (2001) Effects of the polyubiquitin gene Ubi. U4 leader intron and first ubiquitin monomer on reporter gene expression in *Nicotiana tabacum*. *Plant Mol Biol.* 45: 655-667.
- Postma J, Bonants PJ & Van Os EA** (2001) Population dynamics of *Pythium aphanidermatum* in cucumber grown in closed systems. *Meded Fak Landbouwkd Toegep Biol Wet.* 66: 47-59.
- Powell J** (1998) Enhanced concatemer cloning-a modification to the SAGE (Serial Analysis of Gene Expression) technique. *Nucleic Acids Res.* 26: 3445-3446.
- Prat A, Schmid SR, Buser P, Blum S, Trachsel H, Nielsen PJ & Linder P** (1990) Expression of translation initiation factor 4A from yeast and mouse in *Saccharomyces cerevisiae*. *Biochim Biophys Acta.* 1050: 140-145.
- Pruitt RE & Meyerowitz EM** (1986) Characterization of the genome of *Arabidopsis thaliana*. *J Mol Biol.* 187: 169-183.
- Pufky J, Qiu Y, Rao MV, Hurban P & Jones AM** (2003) The auxin-induced transcriptome for etiolated *Arabidopsis* seedlings using a structure/function approach. *Funct Integr Genomics.* 3: 135-143. Epub 2003 Nov 2026.
- Puthoff DP, Nettleton D, Rodermeil SR & Baum TJ** (2003) *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. *Plant J.* 33: 911-921.
- Raes J, Rohde A, Christensen JH, Van De Peer Y & Boerjan W** (2003a) Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol.* 133: 1051-1071.
- Raes J, Vandepoele K, Simillion C, Saeys Y & Van de Peer Y** (2003b) Investigating ancient duplication events in the *Arabidopsis* genome. *J Struct Funct Genomics.* 3: 117-129.
- Raeymaekers L** (1993) Quantitative PCR: theoretical considerations with practical implications. *Anal Biochem.* 214: 582-585.
- Rajeevan MS, Vernon SD, Taysavang N & Unger ER** (2001) Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. *J Mol Diagn.* 3: 26-31.
- Ramonell KM, Zhang B, Ewing RM, Chen Y, Xu D, Stacey G & Somerville S** (2002) Microarray analysis of chitin elicitation in *Arabidopsis thaliana*. *Mol Plant Pathol.* 3: 301-311.
- Rashotte AM, Carson SD, To JP & Kieber JJ** (2003) Expression profiling of cytokinin action in *Arabidopsis*. *Plant Physiol.* 132: 1998-2011.
- Ray BK, Lawson TG, Kramer JC, Cladaras MH, Grifo JA, Abramson RD, Merrick WC & Thach RE** (1985) ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem.* 260: 7651-7658.
- Regenberg B, During-Olsen L, Kielland-Brandt MC & Holmberg S** (1999) Substrate specificity and gene expression of the amino-acid permeases in *Saccharomyces cerevisiae*. *Curr Genet.* 36: 317-328.
- Reintanz B, Szyroki A, Ivashikina N, Ache P, Godde M, Becker D, Palme K & Hedrich R** (2002) AtKC1, a silent *Arabidopsis* potassium channel alpha-subunit modulates root hair K<sup>+</sup> influx. *Proc Natl Acad Sci U S A.* 99: 4079-4084.
- Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M, Weissman S & Snyder M** (2003) The transcriptional activity of human chromosome 22. *Genes Dev.* 17: 529-540.
- Rizhsky L, Liang H & Mittler R** (2003) The water-water cycle is essential for chloroplast protection in the absence of stress. *J Biol Chem.* 278: 38921-38925.
- Rogers GW, Lima WF & Merrick WC** (2001a) Further characterization of the helicase activity of eIF4A - Substrate specificity. *J Biol Chem.* 276: 12598-12608.
- Rogers GW, Richter NJ, Lima WF & Merrick WC** (2001b) Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J Biol Chem.* 276: 30914-30922.
- Rombauts S, Florquin K, Lescot M, Marchal K, Rouze P & van de Peer Y** (2003) Computational approaches to identify promoters and cis-regulatory elements in plant genomes. *Plant Physiol.* 132: 1162-1176.
- Roussel DL & Bennett KL** (1993) glh-1, a germ-line putative RNA helicase from *Caenorhabditis*, has four zinc fingers. *Proc Natl Acad Sci U S A.* 90: 9300-9304.
- Ruan Y, Gilmore J & Conner T** (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J.* 15: 821-833.
- Rubin GM, Hong L, Brokstein P, Evans-Holm M, Frise E, Stapleton M & Harvey DA** (2000) A *Drosophila* complementary DNA resource. *Science.* 287: 2222-2224.

- Ruuska SA, Girke T, Benning C & Ohlrogge JB** (2002) Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell*. 14: 1191-1206.
- Rychlik W & Rhoads RE** (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.* 17: 8543-8551.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB & Erlich HA** (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 239: 487-491.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA & Arnheim N** (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 230: 1350-1354.
- Sakai H, Medrano LJ & Meyerowitz EM** (1995) Role of SUPERMAN in maintaining *Arabidopsis* floral whorl boundaries. *Nature*. 378: 199-203.
- Sambrook J, Fritsch EF & Maniatis T** (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, New York.
- Samson F, Brunaud V, Balzergue S, Dubreucq B, Lepiniec L, Pelletier G, Caboche M & Lecharny A** (2002) FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Res.* 30: 94-97.
- Sankoff D** (2001) Gene and genome duplication. *Curr Opin Genet Dev.* 11: 681-684.
- Sawant SV, Kiran K, Singh PK & Tuli R** (2001) Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants. *Plant Physiol.* 126: 1630-1636.
- Sawbridge T, Ong E-K, Binnion C, Emmerling M, Meath K, Nunan K, O'Neill M, O'Toole F, Simmonds J & Wearne K** (2003) Generation and analysis of expressed sequence tags in white clover (*Trifolium repens* L.). *Plant Science*. 165: 1077-1087.
- Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M & Wisman E** (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. *Plant Cell*. 13: 113-123.
- Schena M, Shalon D, Davis RW & Brown PO** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 270: 467-470.
- Schenk PM, Kazan K, Manners JM, Anderson JP, Simpson RS, Wilson IW, Somerville SC & Maclean DJ** (2003) Systemic gene expression in *Arabidopsis* during an incompatible interaction with *Alternaria brassicicola*. *Plant Physiol.* 132: 999-1010.
- Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC & Manners JM** (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc Nat Acad Sci U S A*. 97: 11655-11660.
- Schmid SR & Linder P** (1991) Translation initiation factor 4A from *Saccharomyces cerevisiae*: analysis of residues conserved in the D-E-A-D family of RNA helicases. *Mol Cell Biol.* 11: 3463-3471.
- Schmid SR & Linder P** (1992) D-E-A-D protein family of putative RNA helicases. *Mol Microbiol.* 6: 283-291.
- Schnerr H, Niessen L & Vogel RF** (2001) Real time detection of the *tri5* gene in *Fusarium* species by lightcycler-PCR using SYBR Green I for continuous fluorescence monitoring. *Int J Food Microbiol.* 71: 53-61.
- Schoof H & Karlowski WM** (2003) Comparison of rice and *Arabidopsis* annotation. *Curr Opin Plant Biol.* 6: 106-112.
- Schroeder DF, Gahrtz M, Maxwell BB, Cook RK, Kan JM, Alonso JM, Ecker JR & Chory J** (2002) De-etiolated 1 and damaged DNA binding protein 1 interact to regulate *Arabidopsis* photomorphogenesis. *Curr Biol.* 12: 1462-1472.
- Schwarz-Sommer Z** (1987) The significance of plant transposable elements in biological processes. *Results Probl Cell Differ.* 14: 213-221.
- Seki M, Carninci P, Nishiyama Y, Hayashizaki Y & Shinozaki K** (1998) High-efficiency cloning of *Arabidopsis* full-length cDNA by biotinylated CAP trapper. *Plant J.* 15: 707-720.
- Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y & Shinozaki K** (2002a) Monitoring the expression pattern of around 7,000 *Arabidopsis* genes under ABA treatments using a full-length cDNA microarray. *Funct Integr Genomics.* 2: 282-291.
- Seki M, Narusaka M, Abe H, Kasuga M, Yamaguchi-Shinozaki K, Carninci P, Hayashizaki Y & Shinozaki K** (2001) Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell.* 13: 61-72.
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y & Shinozaki K** (2002b) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279-292.
- Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K,**

- Shinagawa A & Shinozaki K** (2002c) Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science*. 296: 141-145.
- Seki M, Satou M, Sakurai T & Shinozaki K** (2002d) RIKEN *Arabidopsis* full-length cDNA database. *Trends Plant Sci*. 7: 562-563.
- Shimada Y, Goda H, Nakamura A, Takatsuto S, Fujioka S & Yoshida S** (2003) Organ-specific expression of brassinosteroid-biosynthetic genes and distribution of endogenous brassinosteroids in *Arabidopsis*. *Plant Physiol*. 131: 287-297.
- Skaf JS, Schultz MH, Hirata H & de Zoeten GA** (2000) Mutational evidence that the VPg is involved in the replication and not the movement of Pea enation mosaic virus-1. *J Gen Virol*. 81: 1103-1109.
- Smale ST & Kadonaga JT** (2003) The RNA polymerase II core promoter. *Annu Rev Biochem*. 72: 449-479.
- Smith CM & Steitz JA** (1997) Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell*. 89: 669-672.
- Smith RD, Brown B, Ikononi P & Schechter AN** (2003) Exogenous reference RNA for normalization of real-time quantitative PCR. *Biotechniques*. 34: 88-91.
- Smith SW, Overbeek R, Woese CR, Gilbert W & Gillevet PM** (1994) The genetic data environment an expandable GUI for multiple sequence analysis. *Comput Appl Biosci*. 10: 671-675.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D & Futcher B** (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell*. 9: 3273-3297.
- Stapleton M, Carlson J, Brokstein P, Yu C, Champe M, George R, Guarin H, Kronmiller B, Pacleb J, Park S, Wan K, Rubin GM & Celniker SE** (2002a) A *Drosophila* full-length cDNA resource. *Genome Biol*. 3: research0080.0081-0080.0088.
- Stapleton M, Liao G, Brokstein P, Hong L, Carninci P, Shiraki T, Hayashizaki Y, Champe M, Pacleb J, Wan K, Yu C, Carlson J, George R, Celniker S & Rubin GM** (2002b) The *Drosophila* gene collection: identification of putative full-length cDNAs for 70% of *D. melanogaster* genes. *Genome Res*. 12: 1294-1300.
- Starr DB, Hoopes BC & Hawley DK** (1995) DNA bending is an important component of site-specific recognition by the TATA binding protein. *J Mol Biol*. 250: 434-446.
- Stasolla C, van Zyl L, Egertsdotter U, Craig D, Liu WB & Sederoff RR** (2003) The effects of polyethylene glycol on gene expression of developing white spruce somatic embryos. *Plant Physiol*. 131: 49-60.
- Strasser K & Hurt E** (2001) Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature*. 413: 648-652.
- Stubner S** (2002) Enumeration of 16S rDNA of *Desulfotomaculum* lineage 1 in rice field soil by real-time PCR with SybrGreen detection. *J Microbiol Methods*. 50: 155-164.
- Sun CW & Callis J** (1997) Independent modulation of *Arabidopsis thaliana* polyubiquitin mRNAs in different organs and in response to environmental changes. *Plant J*. 11: 1017-1027.
- Suzuki H, Nakayama T, Yonekura-Sakakibara K, Fukui Y, Nakamura N, Yamaguchi MA, Tanaka Y, Kusumi T & Nishino T** (2002) cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme a:anthocyanidin 3-o-glucoside-6- $\alpha$ -malonyltransferase from dahlia flowers. *Plant Physiol*. 130: 2142-2151.
- Svitkin YV, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham GJ & Sonenberg N** (2001) The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA*. 7: 382-394.
- Tan BC, Joseph LM, Deng WT, Liu LJ, Li QB, Cline K & McCarty DR** (2003) Molecular characterization of the *Arabidopsis* 9-cis epoxycarotenoid dioxygenase gene family. *Plant J*. 35: 44-56.
- Tanner NK, Cordin O, Banroques J, Doere M & Linder P** (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell*. 11: 127-138.
- Tanner NK & Linder P** (2001) DEXD/H box RNA helicases: From generic motors to specific dissociation functions. *Mol Cell*. 8: 251-262.
- Tao Y, Xie Z, Chen W, Glazebrook J, Chang HS, Han B, Zhu T, Zou G & Katagiri F** (2003) Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell*. 15: 317-330.
- Tepperman JM, Zhu T, Chang HS, Wang X & Quail PH** (2001) Multiple transcription-factor genes are early targets of phytochrome A signaling. *Proc Nat Acad Sci U S A*. 98: 9437-9442.
- Teramoto H, Nakamori A, Minagawa J & Ono T** (2002) Light-intensity-dependent expression of Lhc gene family encoding light-harvesting chlorophyll-a/b proteins of photosystem II in *Chlamydomonas reinhardtii*. *Plant Physiol*. 130: 325-333.
- Terry CF, Shanahan DJ, Ballam LD, Harris N, McDowell DG & Parkes HC** (2002) Real-time detection of genetically modified soya using Lightcycler and ABI 7700 platforms with TaqMan, Scorpion, and SYBR Green I chemistries. *J AOAC Int*. 85: 938-944.
- Thureau V, Dehais P, Serizet C, Hilsen P, Rouze P & Aubourg S** (2003) Automatic design of gene-specific sequence tags for genome-wide functional studies. *Bioinformatics*. 19: 2191-2198.



- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU & Saedler H** (2000) A short history of MADS-box genes in plants. *Plant Mol Biol.* 42: 115-149.
- Thomas JH** (1993) Thinking about genetic redundancy. *Trends Genet.* 9: 395-399.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG** (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
- Thompson JD, Higgins DG & Gibson TJ** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
- Tollervey D & Kiss T** (1997) Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Biol.* 9: 337-342.
- Tremousaygue D, Garnier L, Bardet C, Dabos P, Herve C & Lescure B** (2003) Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in *Arabidopsis thaliana* cycling cells. *Plant J.* 33: 957-966.
- Tremousaygue D, Manevski A, Bardet C, Lescure N & Lescure B** (1999) Plant interstitial telomere motifs participate in the control of gene expression in root meristems. *Plant J.* 20: 553-561.
- Ulm R, Baumann A, Oravec A, Mate Z, Adam E, Oakeley EJ, Schafer E & Nagy F** (2004) Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of *Arabidopsis*. *Proc Natl Acad Sci U S A.* 101: 1397-1402.
- Vaitilingom M, Pijnenburg H, Gendre F & Brignon P** (1999) Real-time quantitative PCR detection of genetically modified Maximizer maize and Roundup Ready soybean in some representative foods. *J Agric Food Chem.* 47: 5261-5266.
- van Helden J, Andre B & Collado-Vides J** (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol.* 281: 827-842.
- van Noort V, Snel B & Huynen MA** (2003) Predicting gene function by conserved co-expression. *Trends Genet.* 19: 238-242.
- Vandepoele K, Raes J, De Veylder L, Rouze P, Rombauts S & Inze D** (2002) Genome-wide analysis of core cell cycle genes in *Arabidopsis*. *Plant Cell.* 14: 903-916.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A & Speleman F** (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3: research0034.0031-research0034.0011.
- Velculescu VE, Vogelstein B & Kinzler KW** (2000) Analysing uncharted transcriptomes with SAGE. *Trends Genet.* 16: 423-425.
- Velculescu VE, Zhang L, Vogelstein B & Kinzler KW** (1995) Serial analysis of gene expression. *Science.* 270: 484-487.
- Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Jr., Hieter P, Vogelstein B & Kinzler KW** (1997) Characterization of the yeast transcriptome. *Cell.* 88: 243-251.
- Vision TJ, Brown DG & Tanksley SD** (2000) The origins of genomic duplications in *Arabidopsis*. *Science.* 290: 2114-2117.
- Volkov RA, Panchuk, II & Schoffl F** (2003) Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR. *J Exp Bot.* 54: 2343-2349.
- Wang R, Okamoto M, Xing X & Crawford NM** (2003a) Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* 132: 556-567.
- Wang RC, Guegler K, LaBrie ST & Crawford NM** (2000a) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. *Plant Cell.* 12: 1491-1509.
- Wang RC, Okamoto M, Xing XJ & Crawford NM** (2003b) Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol.* 132: 556-567.
- Wang SM & Rowley JD** (1998) A strategy for genome-wide gene analysis: integrated procedure for gene identification. *Proc Natl Acad Sci U S A.* 95: 11909-11914.
- Wang Y, Duby G, Purnelle B & Boutry M** (2000b) Tobacco VDL gene encodes a plastid DEAD box RNA helicase and is involved in chloroplast differentiation and plant morphogenesis. *Plant Cell.* 12: 2129-2142.
- Wang Z, Taramino G, Yang D, Liu G, Tingey SV, Miao GH & Wang GL** (2001) Rice ESTs with disease-resistance gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. *Mol Genet Genomics.* 265: 302-310.
- Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Shownkeen R, Halloran N & et al.** (1992) A survey of expressed genes in *Caenorhabditis elegans*. *Nat Genet.* 1: 114-123.

- Weaver PL, Sun C & Chang TH** (1997) Dbp3p, a putative RNA helicase in *Saccharomyces cerevisiae*, is required for efficient pre-rRNA processing predominantly at site A3. *Mol Cell Biol.* 17: 1354-1365.
- Weller SA, Elphinstone JG, Smith NC, Boonham N & Stead DE** (2000) Detection of *Ralstonia solanacearum* strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. *Appl Environ Microbiol.* 66: 2853-2858.
- Wendel JF** (2000) Genome evolution in polyploids. *Plant Mol Biol.* 42: 225-249.
- White JA, Todd J, Newman T, Focks N, Girke T, de Hauduy OM, Jaworski JG, Ohlrogge JB & Benning C** (2000) A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol.* 124: 1582-1594.
- Whitham SA, Quan S, Chang HS, Cooper B, Estes B, Zhu T, Wang X & Hou YM** (2003) Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *Plant J.* 33: 271-283.
- William DA, Su Y, Smith MR, Lu M, Baldwin DA & Wagner D** (2004) Genomic identification of direct target genes of LEAFY. *Proc Natl Acad Sci U S A.* 101: 1775-1780.
- Wiseman G** (2002) State of the art and limitations of quantitative polymerase chain reaction. *J AOAC Int.* 85: 792-796.
- Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA & Balis UJ** (1997) The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques.* 22: 176-181.
- Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T & Sasaki T** (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell.* 14: 525-535.
- Wurbach E, Yuen T & Sealfon S** (2003) Focused microarray analysis. *Methods.* 31: 306-316.
- Xu WY, Bak S, Decker A, Paquette SM, Feyereisen R & Galbraith DW** (2001) Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene.* 272: 61-74.
- Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A & Ecker JR** (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science.* 302: 842-846.
- Yamamoto K & Sasaki T** (1997) Large-scale EST sequencing in rice. *Plant Mol Biol.* 35: 135-144.
- Yang JM, Park S, Kamdem DP, Keathley DE, Retzel E, Paule C, Kapur V & Han KH** (2003) Novel gene expression profiles define the metabolic and physiological processes characteristic of wood and its extractive formation in a hardwood tree species, *Robinia pseudoacacia*. *Plant Mol Biol.* 52: 935-956.
- Yazaki J, Kishimoto N, Nakamura K, Fujii F, Shimbo K, Otsuka Y, Wu J, Yamamoto K, Sakata K, Sasaki T & Kikuchi S** (2000) Embarking on rice functional genomics via cDNA microarray: use of 3' UTR probes for specific gene expression analysis. *DNA Res.* 7: 367-370.
- Yean D & Gralla J** (1997) Transcription reinitiation rate: a special role for the TATA box. *Mol Cell Biol.* 17: 3809-3816.
- Yokoyama R & Nishitani K** (2001) A comprehensive expression analysis of all members of a gene family encoding cell-wall enzymes allowed us to predict cis-regulatory regions involved in cell-wall construction in specific organs of *Arabidopsis*. *Plant Cell Physiol.* 42: 1025-1033.
- Young ND** (2000) The genetic architecture of resistance. *Curr Opin Plant Biol.* 3: 285-290.
- Yu JP, Nickels R & McIntosh L** (2001) A genome approach to mitochondrial-nuclear communication in *Arabidopsis*. *Plant Physiol Biochem.* 39: 345-353.
- Zegzouti H, Jones B, Frasse P, Marty C, Maitre B, Latch A, Pech JC & Bouzayen M** (1999) Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. *Plant J.* 18: 589-600.
- Zhang J, Zhang YP & Rosenberg HF** (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet.* 30: 411-415.
- Zhang PF, Chopra S & Peterson T** (2000) A segmental gene duplication generated differentially expressed myb-homologous genes in maize. *Plant Cell.* 12: 2311-2322.
- Zhang XS & Choi JH** (2001) Molecular evolution of calmodulin-like domain protein kinases (CDPKs) in plants and protists. *J Mol Evol.* 53: 214-224.

- Zhao J, Jin SB, Bjorkroth B, Wieslander L & Daneholt B** (2002) The mRNA export factor Dbp5 is associated with Balbiani ring mRNP from gene to cytoplasm. *EMBO J.* 21: 1177-1187.
- Zhao YF, Thilmony R, Bender CL, Schaller A, He SY & Howe GA** (2003) Virulence systems of *Pseudomonas syringae* pv. tomato promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. *Plant J.* 36: 485-499.
- Zheng ZL, Nafisi M, Tam A, Li H, Crowell DN, Chary SN, Schroeder JI, Shen JJ & Yang ZB** (2002) Plasma membrane-associated ROP10 small GTPase is a specific negative regulator of abscisic acid responses in *Arabidopsis*. *Plant Cell.* 14: 2787-2797.
- Zhu T, Budworth P, Han B, Brown D, Chang HS, Zou GZ & Wang X** (2001) Toward elucidating the global gene expression patterns of developing *Arabidopsis*: Parallel analysis of 8 300 genes by a high-density oligonucleotide probe array. *Plant Physiol Biochem.* 39: 221-242.
- Zik M & Irish VF** (2003) Global identification of target genes regulated by *APETALA3* and *PISTILLATA* floral homeotic gene action. *Plant Cell.* 15: 207-222.

## Annexes

- Article 1** **Mingam A, Gachon C and Charrier B.** Quantifying nucleic acids with fluorescent probes: contribution of real-time PCR to applied and fundamental plant research. *Advances in Plant Physiology*. Vol. 7
- Article 2** **Gachon C, Mingam A and Charrier B.** Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany*. Sous presse
- Annexe 3** Code de la macro « Modifie1\_TraitementFichierCSV »

**Quantifying nucleic acids with fluorescent probes:  
Contribution of real-time PCR to applied and fundamental  
plant research.**

*Advances in Plant Physiology*. 2003. Vol. 7, sous presse.

Annaïck Mingam, Claire Gachon, Bénédicte Charrier

Institut de Biotechnologie des Plantes UMR CNRS 8618  
Université Paris-Sud  
91405 Orsay cedex. France.

## SUMMARY

## 1.0. WHAT USE FOR REAL-TIME PCR IN PLANT RESEARCH?

**1.1-Looking for a needle in a haystack****1.2-From transcript to function: to what extent is the transcript level informative?**

## 2.0. ORIGINS OF REAL-TIME PCR

**2.1-Polymerase Chain Reaction (PCR)****2.2-Semi-quantitative (RT-) PCR****2.3-Competitive quantitative PCR****2.4-Real-time (RT-) PCR**

## 3.0-FEATURES OF REAL-TIME PCR

**3.1-Detection**

3.1.1-Intercalating agent

3.1.2-Probes based on the Fluorescence Resonance Energy Transfer process (FRET)

**3.2-Rapidity****3.3-Sensitivity****3.4-Specificity**

3.4.1-Specificity during the cDNA synthesis

3.4.2-Specificity during the PCR

3.4.3-Checking the specificity after the PCR

**3.5-Quantification**

3.5.1-Exponential amplification

3.5.2-Choice of a standard molecule

3.5.2.1- Single strand matrix

- Mixture of RNAs

- Specific RNA

3.5.2.2- Double-strand matrix

- cDNA clone

- Free PCR product

- Genomic DNA

**3.6-Presentation of the results**

3.6.1-Absolute level of copies

3.6.2-Percentage

3.6.3-Compared to an endogenous gene

- Constitutive gene

- Single copy gene

3.6.4-Compared to a sample of the same study

## 4.0-APPLICATIONS

**4.1-Detecting and quantifying the presence of foreign plant material.**

4.1.1-Food safety and agricultural concerns: detecting contamination with foreign DNA

-DNA from another wild-type plant

-DNA from genetically modified plants

4.1.2-Estimating transgene copy number in a plant genome

-Multiple insertions at different loci

-State of zygosity of the transgene locus

**4.2-Detecting and quantifying plant parasites or symbiotes.**

4.2.1-Health monitoring of cultivated species

4.2.2-Studies of plant response to pathogen infection

**4.3-Studying the expression profile of multigene families****4.4-Confirmation of large-scale expression studies****4.5- Molecular phenotype of mutant and transgenic plants**

4.5.1- Measuring the effect of a mutation on its target gene

4.5.2- Assessing the impact on other gene expression

**4.6-Miscellaneous**

## 5.0- LIMITATION OF REAL-TIME PCR TECHNIQUES IN PLANT RESEARCH

## 6.0- CONCLUSION

## 1.0- WHAT USE FOR REAL-TIME PCR IN PLANT RESEARCH?

While real-time quantitative PCR had extensively been used in animal and medical sciences for the past ten years (reviewed in Wilhelm and Pingoud, 2003), an explosion of its applications in the domain of plant research was activated only in year 2000. Below is briefly presented why real-time (RT-) PCR has attracted so much interest in the plant community. A whole section reporting the detailed applications of real-time PCR in plant research is presented further in the text.

### 1.1-Looking for a needle in a haystack

The production of genetically modified plants (GMPs) and their sale on the food market, were simultaneously accompanied by a request for a sensitive and reproducible technique able to trace them in a reliable way in different kinds of environments. Indeed, the presence of transgenic pollen or seeds in a non-transgenic field must be detectable, and the percentage of genetically modified food within a processed product be determined precisely. Therefore, food safety agencies, strongly supported by consumer associations, have developed protocols allowing a fast, reliable and reproducible detection and quantification of genetically modified materials in food. A review recently published on this issue (Auer, 2003), presents the different techniques currently used to achieve this aim.

### 1.2- From transcript to function: to what extent is the transcript level informative?

Knowing the expression profile of a gene may be informative with regard to its function. Transcription factors are highly regulated at the transcriptional level, and while post-translational modifications are largely mobilized to regulate enzyme activities, the level of transcripts of some enzyme encoding genes may also reflect the activity of the corresponding enzymes (Suzuki *et al.*, 2002; Svensson *et al.*, 2002). In practice, gene expression profile can be described by either measuring the steady-state level of transcripts (Northern blot, reverse-transcriptase PCR), or measuring the activity of its promoter, using one of the following reporter systems, *i.e.*  $\beta$ -glucuronidase (GUS), luciferase or green-fluorescent protein GFP, or a combination of them (de Ruijter *et al.*, 2003). However, these reporter systems only measure the transcriptional activity, and before presuming that the data reflect the level of steady-state transcripts, the rate of transcript degradation has to be considered as negligible. On the other hand, compared to Northern, reverse-transcriptase PCR (RT-PCR) requires additional steps, such as reverse-transcription, which putatively result in technical biases. Nevertheless, several authors reported the confirmation of real-time RT-PCR data by Northern blot analysis (Brown *et al.*, 2003; Schenk *et al.*, 2003).

## 2.0- ORIGINS OF REAL-TIME PCR

### 2.1-Polymerase Chain Reaction (PCR)

Real-time PCR emerged from the intensive use of the “Polymerisation Chain Reaction” (PCR). The latter was developed by Saiki and collaborators (1985), Mullis *et al.* (1986) and Mullis and Faloona (1987). Originally, the reaction was catalysed by the Klenow fragment of *Echerischia coli* DNA polymerase I. Later, the original protocol was improved by replacing this thermo-sensitive DNA polymerase with a thermo-resistant enzyme, the Taq Polymerase

extracted from the thermo-resistant bacterium *Thermophyllus aquaticus* (Chien *et al.*, 1976). This major discovery allowed Kary B. Millis to be awarded the Nobel Prize of Chemistry in 1993. Indeed, before 1985, obtaining a large number of copies of identical DNA required first a cloning step into a plasmid, and then a multiplication step after introducing the recombinant plasmid into a fast-growing bacterium. Extraction of the plasmid was an additional step. Thanks to PCR, the time required for generating large quantities of a specific DNA fragment was reduced from several days to a few hours.

At the beginning of the PCR reaction, the amount of DNA doubles at each cycle, so that the total amount of DNA molecules is  $2^n \times Q_0$ , where  $Q_0$  is the initial matrix quantity and  $n$  the cycle number. Thus, a 20-cycle reaction commonly yields a  $10^6$ -fold amplification of the initial product. However, due to the progressive exhaustion of reagents during the PCR run, the amplification yield decreases in the last cycles, and reaches a plateau as illustrated in figure 1. For this reason, the initial amount of DNA cannot be reliably inferred from the amount of amplified DNA.

A review on the use of PCR and its modified versions was published by Taylor and Robinson (1998).

## 2.2-Semi-quantitative (RT-) PCR

PCR may also be used for the quantification of a specific mRNA. For this purpose, a preliminary step is required in order to convert the mRNA into the corresponding cDNA. Indeed, the Taq DNA polymerase is DNA-dependent and cannot use RNA as a matrix. This additional step is performed using the reverse-transcriptase enzyme, an RNA-dependent DNA polymerase, which requires an oligonucleotide to initiate the polymerisation. This reverse-transcriptase PCR (RT-PCR) protocol was first developed to study relative levels of specific transcripts in different growth conditions (Rappolee *et al.*, 1988) and was routinely used in many laboratories since then.

However, the quantification of the PCR product should be performed before reaching the saturation phase in order to accurately compare the initial amount of DNA in different samples. In practice, aliquots are taken at different steps of the PCR reaction, so that the earliest ones correspond to the exponential phase of the PCR amplification. After running the amplified fragments on an agarose gel in the presence of a fluorescent DNA-intercalating agent, the intensity of the bands is quantified under UV, using a quantification software. Then, an amplification curve is drawn from the measurements of the amplified fragments, and only aliquots corresponding to the exponential phase are used for the comparison between the different samples. In addition, in order to check the specificity of the PCR reaction, DNA must be transferred to a Hybond membrane, and hybridised with a labelled probe corresponding to the amplified fragment. Technical features and advantages of this method were initially reviewed by Ferre (1992).

## 2.3-Competitive quantitative PCR

Different sophistications of semi-quantitative PCR have been introduced and are reviewed by Raeymaekers (1993). In competitive PCR, a standard DNA of known concentration is added to the reaction mixture before starting the PCR. This standard should be amplifiable with the same efficiency as the target present in the samples. Both samples and standard aliquots are taken at different steps of the PCR run, and after migration on an agarose gel, the intensity of the bands is compared to that of the standard. When the intensities of the bands are identical, then one considers that sample and standard had the same amount of DNA at the beginning of



the reaction (Becker-Andre and Hahlbrock, 1989; Gilliland *et al.*, 1990). The so-called non-competitive RT-PCR proceeds the same way, except that the quantification occurs before the two targets are competitive (Freeman *et al.*, 1999). Both techniques bring valuable results, but they are time consuming and require to open PCR tubes each time an aliquot is collected, which may lead to artefacts, such as sample contamination or water loss. Alternatively, one might prepare as many tubes as the number of aliquots to be collected, which is even more tedious and expensive. Therefore, in the beginning of the 90s, methods were set up to record the amplified product in real-time, without interfering with the PCR process.

## 2.4-Real-time (RT-) PCR

The first PCR machine able to record DNA amplification in real-time was described in 1993 by Higuchi and collaborators (Higuchi *et al.*, 1993). The device consisted in lighting up all the tubes of a PCR machine simultaneously with a UV light. The DNA-intercalating agent ethidium bromide, fluorescing when trapped between double-stranded DNA, was added to the reaction mixture. The light emitted by this agent in response to UV irradiation was measured and recorded by a video device during the elongation step of each PCR cycle. The authors showed that the intensity of the fluorescence was directly proportional to the amount of PCR products in each tube, provided that the amplified fragment was small. Figure 1 illustrates how this improvement allowed to calculate the amount of amplified product during the exponential phase of the PCR reaction.

From the fluorescence signal recorded in real-time, an amplification curve is drawn for each sample. A theoretical fluorescence threshold is then defined in the exponential phase of the PCR reaction. Each sample requires a number of cycles to reach this threshold (called Ct or CP depending on the PCR system). This value is then used to accurately compare the initial amount of DNA present in two samples (for example, control versus stressed material), following equation calculations proposed by real-time machine manufacturers, or according to Pfaffl (2001). Alternatively, the Ct or CP value can be used to calculate the absolute amount of initial DNA in each sample, thanks to the comparison with a reference dilution series of known copy number, as illustrated in Charrier *et al.* (2002).

Higuchi *et al.* (1993) illustrated the technical advantages of this technique by performing the quantification of  $10^2$  to  $10^9$  molecules among 300 nanograms of human genomic DNA, in a reliable and reproducible way. Three years later, Heid and collaborators (1996) proposed to measure the amount of the PCR product using a fluorogenic oligonucleotide, called the TaqMan probe, in addition to the two specific oligonucleotides usually required for every PCR reaction. The TaqMan probe is labelled by both a fluorochrome and a quencher, and is specific for the DNA target. When it hybridises to the DNA target, the quencher prevents the emission of light. Along the polymerisation step, the 5'→3' exonuclease activity of the Taq DNA polymerase results in the degradation of this oligonucleotide. The quencher is then released in the reaction mixture, and the fluorochrome emits light, which is simultaneously recorded by the video system. This method allows to increase the signal specificity. Indeed, on the one hand, three oligonucleotides are needed, and on the other hand, the fluorescence is emitted only when the specific product is being amplified. Furthermore, using different TaqMan probes allows the simultaneous amplification of different specific targets (referred to as multiplex PCR), provided the TaqMan probes are labelled with fluorochromes with different wavelength features.

The LightCycler™ was first commercialised in 1997 as another type of real-time PCR machine (Wittwer *et al.*, 1997a; 1997b). Each reaction mixture is introduced into one capillary, and the heating process is performed by light. The length of the PCR reaction is

tremendously reduced. The authors also used SYBR Green® as a new intercalating agent. Excitation of SYBR Green® is performed at 485 nm by means of a halogen lamp. The intensity of the emitted light at 520 nm (green light) is captured by a CCD camera, and stored on the hard disk of the computer. Since SYBR Green® emits a residual fluorescence even when it is not associated to double stranded DNA, a sample containing the reaction mixture and devoid of nucleic acids should always be included as a negative control, and the measured fluorescence subtracted from the sample values. In practice, SYBR Green® is most commonly purchased in a mastermix containing all other PCR reagents, but it can also be purchased independently, and subsequently added to a home-made mixture, as described by Karsai *et al.* (2002).

Whilst SYBR Green® and TaqMan detection are still predominant, a variety of alternative probes can be purchased from biotechnology companies. All of them rely on FRET, with the fluorochrome and the quencher present on two separate oligonucleotides (hybridisation probes), or on the same oligonucleotide (Molecular Beacons, Scorpion probe). Because up to four specific oligonucleotides may be used (hybridisation probes), these methods ensure a very high specificity. So far, their use was reported in only a handful of plant biology studies.

The fundamentals of real-time PCR are detailed in Bustin (2000; 2002). The machines currently available on the markets, as well as the algorithms necessary for quantification of the PCR product, are presented and discussed by Wilhelm and Pingoud (2003).

Real-time PCR presents a series of characteristics, which make it a very useful and handy technique. In the next section, we focus on the features of real-time PCR, which are specifically relevant to plant research.

### 3.0- FEATURES OF REAL-TIME PCR

#### 3.1-Detection

In the previous section, we rapidly presented the different types of fluorogenic agents used to detect DNA or RNA in real-time along PCR amplification. Here, we discuss the relevance of each of them for plant studies.

##### 3.1.1-Intercalating agent

Because of its intercalating property, SYBR Green® is not sequence specific, and hence can be used for any gene, resulting in a financial advantage for a laboratory testing a large number of genes. Typically, plant studies using SYBR Green® are those analysing the transcript level of a high number of genes involved in the same given process (Anterola *et al.*, 2002; Schenk *et al.*, 2003; Shimada *et al.*, 2003), or members of gene families (Charrier *et al.*, 2002; Orsel *et al.*, 2002; Balbi and Lomax, 2003; Jakab *et al.*, 2003; Kursteiner *et al.*, 2003; Mladek *et al.*, 2003). However, as SYBR Green® is incorporated in non-specific sequences with the same efficiency as specific sequences, the signal measured at each cycle might correspond to amplification of both sequences, thereby compromising the accurate quantification of the specific one.

##### 3.1.2-Probes based on the Fluorescence Resonance Energy Transfer process (FRET)

- Taqman

In practice, TaqMan probes are expensive, and, compared to SYBR Green®, their use is worth if specificity during the amplification process is a critical parameter (see also the section 3.4-Specificity). TaqMan probes are also financially worth when a high number of repetitions of the same experiment must be performed. Typically, they are used in studies aiming at detecting the presence of transgenic DNA in plants intended for commercial purpose, such as events in maize (Hernandez *et al.*, 2003b), peanut and soybean (Schmidt and Parrott, 2001). Alternatively, TaqMan probe can be used for pathogen detection in soil, as for the nematophageous fungus *Plectosphaerella*, a biological control agent of potato cultures (Atkins *et al.*, 2003), or in plant organs, as for the detection of *Ralstonia* in potato tubers (Weller *et al.*, 2000). It is worth noting that TaqMan probes are more frequently used in studies performed for the agribusiness. Nevertheless, they may also be chosen for fundamental studies, as reported in the analysis of the expression profile of some brassinosteroid biosynthesis genes in different organs of *Arabidopsis thaliana* (Shimada *et al.*, 2003), or in the study of the MinE chloroplastic protein encoding gene in *A. thaliana* (Itoh *et al.*, 2001).

Interestingly, Shimada *et al.* (2003) showed that the specificity of real-time PCR was at least 10 times higher when using a TaqMan probe than when using SYBR Green®.

#### - Other probes

The use of other probes, such as hybridisation probes, Molecular Beacons or Scorpion probes, remains very rare in plant. They should be limited to studies requiring an extremely high specificity. Busch *et al.* (2002) nicely illustrated the use of hybridisation probes to quantify the transcript amount of five tobacco genes involved in carotenoid biosynthesis. Likewise, Goncalves *et al.* (2002) used Molecular Beacons to quantify the amount of the sugarcane yellow leaf virus, with a detection sensitivity of 100 femtograms of purified virus. Because of its high specificity, the protocol was proposed as a diagnostic method for the detection of sugarcane viruses. Schena *et al.* (2002) used nested Scorpion PCR as the most specific and sensitive method to detect the presence of pathogen *Rosellinia necatrix* directly from soil.

### 3.2-Rapidity

Compared to non real-time PCR, one of the main advantage of real-time PCR is its rapidity to provide reliable data. As previously mentioned, recording the amplification in real-time avoids collecting samples at different steps of the PCR experiment, making the process much less tedious and time-consuming. In addition, a complete cycle remains very short, with annealing and elongation temperatures identical in some machines. Furthermore, the length of the PCR product is usually very small, *i.e.* < 200 bp, resulting in a very short elongation time. The time technically required to shift temperatures between PCR steps is mainly responsible for the time required to complete the whole PCR experiment. The light cycler is a system using capillaries instead of tubes, which are heated by light instead of a heating block. The time necessary to heat the PCR mixture is then considerably reduced (from 15 sec to 1-2 sec). Typically, the time of a whole PCR run ranges between 20 minutes and 2 hours. In addition, a high number of samples can be tested in the same time (up to 384, Wurmbach *et al.*, 2003). In practice, this feature may be a determining parameter for some studies requiring a high number of repetitions.

In plant research, the detection of pathogens (bacteria *Agrobacterium*, fungus *Fusarium* or mycorrhize *Glomus*) in the root mat of plants grown in soil may be very tedious, as soil must be manipulated, filtered, and purified for inoculation on petri dishes. The gain of time with the

real-time PCR technique was tremendous, since it allowed, in few hours or minutes, both testing directly soil samples for the presence of pathogens DNA, and quantifying the number of cells of pathogen present per ml of soil (Weller and Stead, 2002; Filion *et al.*, 2003a). The use of a portable real-time PCR cyclers in the grapevine of California allowed to screen for grapes infected by the pathogen *Xylella fastidiosa* in one hour on site, thereby short-circuiting the usual diagnosis of Pierce's disease based on leaf colour, and made possible only too late in the season to prevent the propagation of the pathogen (Schaad *et al.*, 2002).

### 3.3- Sensitivity

Because of the PCR amplification step, real-time PCR is a very sensitive technique. Indeed, in their study on the expression of the magnesium chelatase genes in *Chlamydomonas*, Lake and Willows (2003) reported that, while leading to comparable results, real-time RT-PCR requires 10 to 100 less RNA than a Northern blot.

However, the sensitivity of real-time PCR may be discussed, especially when the procedure for sample collection is not optimal. Higgins *et al.* (2003) reported that surface swabs collected during the bioterrorism attack in the United States of America in October 2001, and investigated for the presence of the Anthrax bacteria, were all negative when tested by real-time PCR, whereas cultures made from these swabs managed to reveal their presence. This failure may actually result from a defect in the protocol, as DNA extraction seemed to have been an additional step. Likewise, Spackman *et al.* (2003) reported that real-time PCR, while more rapid, was not more sensitive than the other tests usually used for the detection of influenza viruses.

In studies aiming at finding genes involved in fungal pathogen attack in plants, Schenk *et al.* (2003) noticed that the variation in the transcript level of specific genes involved in the response to pathogen infection, was lower when using real-time RT-PCR than when using cDNA-based micro-array experiments. The different specificities of the two techniques (see next section) may easily explain the difference.

### 3.4-Specificity

In contrast to techniques requiring hybridisation of nucleic acid several hundreds base pair long, such as micro-array and Northern blot, short oligonucleotide-mediated real-time PCR allows a high specificity in the detection of the target sequence. Indeed, in their study of the four pea thioredoxin TRXh isoforms encoding genes, Montrichard *et al.* (2003) noticed that Northern blot gave stronger signals than real-time RT-PCR. They found an explanation in a putative cross-hybridisation of each probe to another isoform of this enzyme. Therefore, Brown *et al.* (2003), after using Northern blot for initial expression analyses, and in a concern of accuracy, used real-time RT-PCR to distinguish the expression of the *Arabidopsis* pathogen-responsive *PDF1-2* gene from three other homologues.

Similar processes occur in micro-array experiment, explaining that real-time PCR often gives different variation ratio than long-fragment hybridisation-based techniques (Schenk *et al.*, 2003).

#### 3.4.1-Specificity during the cDNA synthesis step

Different positions on the RNA sequence can be chosen as the initiation site for cDNA synthesis. In order to obtain a cDNA for every molecule of a pool of total RNAs, corresponding to both ribosomal and messenger RNAs, random hexamers are chosen as priming oligonucleotides (Hammond *et al.*, 2003). However, more typically, oligonucleotides

corresponding to an oligomer (12-18-mer) of thymidine are used, which anneal on the polyadenylated tail of messenger RNAs. This leads to the synthesis of cDNA only for messenger RNAs. Alternatively, and in order to increase the specificity of the reaction, an oligonucleotide specific for each target gene can be used, as illustrated by the study of the expression profile of the very identical 33 members of the XTH gene family in *A. thaliana* (Yokoyama and Nishitani, 2001).

#### 3.4.2-Specificity during the PCR step

The position of the PCR oligonucleotides on the cDNA is an important parameter to take into account. First of all, their position in relation to the position of the oligonucleotides used for the cDNA synthesis has to be considered. Indeed, the further from the reverse-transcription initiation site the PCR oligonucleotides are designed, the less reliable the PCR is, because short cDNAs will be statistically more abundant than long cDNAs. Secondly, the position of the oligonucleotides should be chosen so that the amplified sequence is unique. This is a particularly critical point for studies of plant multigene families. Plant genomes have many more gene families than animal genomes, and in *Arabidopsis thaliana*, two third of the genes are coded by multigene family, with half of them comprising at least five members (AGI, 2000). Therefore, the genomic relationship between the gene of interest and the putative other gene family members should be verified before designing oligonucleotides for real-time PCR. The expression profile of the 10 members of the very conserved gene family coding for the GSK3/Shaggy kinase was analysed using oligonucleotides designed on the 3'UTR of the RNAs, because the percentage of sequence identity in this region was low, compared to the coding sequence (Charrier *et al.*, 2002). Likewise, Busch *et al.* (2002) used primers designed in the C-terminal region of five tobacco genes involved in carotenoid biosynthesis. In order to ultimately verify that the oligonucleotides are specific, a cross-amplification test can be performed. In practice, the designed oligonucleotides are used with an unspecific target, and the amount of the PCR product potentially amplified is considered in the downstream analysis (Busch *et al.*, 2002). Considine *et al.* (2001) noticed that the couple of oligonucleotides designed to specifically amplify only one member of the mango alternative oxidase (AOX) encoding gene family was also able to amplify the other members, yet at a lower rate. They further corrected the expression data in accordance.

#### 3.4.3-Checking the specificity after the PCR reaction

In studies using SYBR Green® as the detection system, the calculation of the melting point ( $T_m$ ) of the PCR product can be carried out at the end of the PCR experiment. In practice, the samples containing the PCR product are heated to 95°C. Along this heating process, the double stranded DNA is denatured, and SYBR Green®, previously trapped within the double stranded DNA, is released, resulting in the fluorescence decreasing. As each nucleic molecule displays a specific  $T_m$ , the dissociation curve allows to detect each different product present in the tube (Halterman *et al.*, 2003; Hammond *et al.*, 2003; Murcha *et al.*, 2003). Pietila *et al.* (2000) took advantage of this option to discriminate between two different species of bacteria. The presence of a single PCR product does not guarantee the identity of this product. Therefore, it is also necessary to sequence the PCR product, to ensure that it corresponds to the gene of interest (Charrier *et al.*, 2002; Halterman *et al.*, 2003).

### 3.5- Quantification

Real-time PCR technique presents the advantage of tolerating up to 7 orders of magnitude for the quantification of PCR products, as originally illustrated by Higuchi *et al.* (1993), Heid *et al.* (1996) and more recently in plants by Charrier *et al.* (2002), Filon *et al.* (2003a) and Schubert *et al.* (2003). However, several conditions must be respected for this technical feature to be reached.

### 3.5.1-Exponential amplification

The accurate quantification of the target molecules present in the sample relies on its exponential PCR amplification, as illustrated in figure 1. Therefore, efficient hybridisation of both oligonucleotides at each PCR cycle must be guaranteed. Different softwares allow to check the absence of dimer or stable tri-dimensional structures within oligonucleotides. In addition, unspecific hybridisation of the oligonucleotides should be avoided, as it will interfere with the efficiency of the PCR.

### 3.5.2-Choice of a standard molecule for the absolute quantification

The quantification is based on a standard curve. Serial dilutions are made with a standard molecule, which is amplifiable with the same efficiency as the target molecule present in the sample. For studies aiming at quantifying the number of transcripts, specific *in vitro* transcribed RNA can be used. In that case, RNA will undergo the same forward steps as the RNA present in the sample (reverse-transcription, PCR). Alternatively, for both RNA and DNA quantification, the use of DNA standard molecules is possible. It can be PCR products specific for the target gene, plasmids recombinant for the target gene, cDNA mixture or genomic DNA. Some are single-stranded, other are double-stranded. Some standards are specific for the target gene, other contain a mixture of targets. Specific standards present the disadvantage of being specific for only one gene of interest. Therefore, as many standards as studied genes have to be prepared, which becomes tedious when a high number of genes are to be analysed. In addition, because of independent preparations of the different standards, comparing levels between different genes is error-prone. Instead, using mixtured standards, such as cDNA preparation or genomic DNA, presents the advantage of being suitable for all the genes present in the plant genome, thereby avoiding errors between several independent standard dilution series.

In any case, the standard DNA must be precisely quantified, so that the mass of the target PCR product can be accurately calculated. Its level will then be expressed in mass units. Alternatively, provided the molecular mass of the standard molecule is known, the number of copies can be determined.

#### 3.5.2.1-Single strand matrix

##### - Mixture of RNAs

RNAs can be used for the preparation of a standard curve. RNAs are reverse-transcribed at the same time as the other samples, and serial dilutions of this cDNA mixture are used to make the standard curve. In order to measure the transcript level of HAK potassium transporters encoding genes in rice, Banuelos *et al.* (2002) used, as a standard matrix, a pool of cDNAs prepared from K<sup>+</sup>-starved roots known to contain a high concentration of HAK transcripts. Hammond *et al.* (2003) exploited the molecular complexity of the cDNA mixture to use it as a single standard for all the genes displaying similar transcriptional behaviour. For

example, a cDNA sample corresponding to RNA extracted from plants starved by phosphate for 100 hours was used as a standard for all the genes displaying a pick of expression 100 hours after starvation.

#### - Specific RNA

The use of *in vitro* transcribed RNA as a standard molecule is rare in plant studies. Yet, Busch *et al.* (2002) reported the use of *in vitro* transcribed RNA specific for five genes of the tobacco carotenoid biosynthetic pathway. These RNAs underwent serial dilutions and were incorporated to the rest of the study in order to be used as standard molecules for transcript quantification.

It is also worth noting that *in vitro* transcribed RNA from plant genes found a utility as an exogenous standard in animal research. Indeed, Smith *et al.*, (2003) reported the use of *in vitro* transcribed RNA of the ribulose biphosphate carboxylase (RUBISCO) for the normalization of the  $\gamma$ -globulin RNA levels in human cells. However, this plant RNA was not used for an absolute quantification, but only for correction of the initial level of transcripts present in the reaction mixture.

#### 3.5.2.2-Double-strand matrix

In case double-stranded DNA is used as a standard for transcript quantification, the number of single-stranded transcript should be doubled compared to the number of PCR product or genome copies.

#### - cDNA or PCR clone

cDNAs corresponding to the gene of interest are often available only when cloned into a vector. Therefore, the quantification of DNA or RNA using such material as a standard were extensively performed, as illustrated for the quantification of potato and tomato DNA in processed food (Hernandez *et al.*, 2003a), or for more fundamental studies, such as the expression analysis of the Men1 gene of *A. thaliana* (Itoh *et al.*, 2001). In the case of transgene detection, the recombinant plasmid will be mixed to genomic DNA prior to the PCR amplification, so that it is in the same PCR conditions as the samples (Song *et al.*, 2002). Montrichard *et al.* (Montrichard *et al.*, 2003) cloned the full-length sequences of the four isoforms of the pea NADPH/NADP-thioredoxin encoding genes into plasmids, and calculated the number of copies of these plasmids per mass of DNA, in order to measure the absolute level of transcripts corresponding to each of these isoforms.

#### - Free PCR product

PCR products corresponding to the fragments amplified during real-time PCR may be used instead of recombinant plasmids. Kursteiner *et al.* (2003) used a PCR fragment of the pyruvate decarboxylase (PCD) for the study of the expression profile of the members of the PCD gene family in *A. thaliana*, and Mladek *et al.* (2003) proceeded the same way for the 15 members of the ARIADNE gene family, putative *A. thaliana* homologues of animal E3-ligases.

#### - Genomic DNA

One copy of genome carries one copy of a unique sequence. Similarly, it carries one copy of all the other unique sequences aimed at being quantified. Therefore, using genomic DNA (gDNA) as a quantification standard is particularly worth when studying several genes in the same experiment and for multiplex PCR. Indeed, it makes the comparison between genes more reliable. In addition, the molecular complexity of genomic DNA being much higher than a plasmid or a PCR product, the error made on the calculation of the mass of gDNA used in the standard dilution range will be much smaller than for PCR products or plasmids.

However, this method requires knowing the mass of the genome of the species studied. As an example, the mass of one copy of genome of *Arabidopsis thaliana* has been estimated to be  $0.30 \pm 0.012$  pg (Arumuganathan and Earle, 1991). Hence, one copy of haploid genome weighs 0.15 pg. Therefore, 150 pg of gDNA in a tube corresponds to 100 copies of double-stranded genome, and thereby 200 copies of single-stranded target sequence. Charrier *et al.* (2002) used *Arabidopsis* gDNA for the quantification of 20 different genes coding for 14 serine-threonine kinases, and 6 marker genes involved in a variety of physiological and developmental processes. They showed that way that the absolute level of transcripts varied between 3 to  $10^6$  copies per nanogram of total RNA.

### 3.6-Presentation of the results

Real-time PCR data may be presented in two major different forms. In a limited number of studies, the absolute level of DNA or RNA copies is provided. Knowing the absolute level of DNA or RNA sequences in a given material is necessary when comparisons of quantity or abundance between two biological systems are needed. But most of the plant studies reporting the use of real-time technology are focused on getting precise and reproducible data on gene or transcript abundance, yet only in a relative way. As the relevance of the results largely depends on the way they are expressed and presented, the next section aims at reviewing the different molecules chosen to normalize the level of the target gene.

#### 3.6.1-Absolute level of copies

The power of real-time PCR resides in its capability to calculate the number of copies of a specific target sequence in a very short time. This requires the use of a standard (see the section 3.5 Quantification), for which the number of copies per nanogram of nucleic acid material is previously calculated. Mladek *et al.* (2003) studied nine ARIADNE genes in *A. thaliana*, and estimated their absolute level of transcript being between 1 and 140 copies per nanogram of total RNA. First, due to the technical limitation of real-time PCR, copy numbers under 10 are likely to be not significant, and to correspond to background signal. Charrier *et al.* (2002) studied the transcript level of 20 genes, whose involvement in plant biology was as diverse as carbon metabolism, response to abiotic stresses, and root, leaf, flower or seed development. While the expression level of these genes ranged from 3 to  $10^6$  copies per nanogram of total RNA, the 10 members of the *A. thaliana* GSK3/Shaggy kinase gene family were expressed at similar levels, with an average of 2000 copies of transcript per ng of total RNAs. In contrast to the GSK3/Shaggy gene family, Yokoyama and Nishitani (2001) noticed that the 33 members of the *Arabidopsis* xyloglucan endotransglycosylase/hydrolase (XTH) gene family displayed very big discrepancy in their transcript levels, with the highest level at 20,000 copies per ng of total RNA, and the lowest at less than 100 copies. Montrichard *et al.* (Montrichard *et al.*, 2003) showed that the pea thioredoxin TRXh transcripts were present in cotyledons, embryos, green leaves and roots with 50,000 to 1.5 millions copies. Absolute transcript levels may also be expressed in moles  $\times$  mass of RNA<sup>-1</sup> (Collakova and DellaPenna, 2003).



In pathogenesis studies, several examples illustrated the use of real-time PCR to quantify the abundance of pathogenic agents around or within the plant (Tierens *et al.*, 2001). Weller *et al.* (2000) reported the detection of 100 cells per ml of culture of the potato pathogen *Ralstonia* with the Taqman technology.

### 3.6.2-Percentage

Expressing the abundance of DNA or RNA sequences within a given material may take into account the abundance of the other sequences analysed in the same study. This is particularly relevant when the latter genes all belong to the same gene family, and this makes even more sense when all the members of the gene family are considered in the study. The abundance of one given gene is then expressed as a percentage of the sum of the levels of all the other members of the gene family. This system of representation presents the advantage of clearly displaying the contribution of each gene to the family. Berger *et al.* (2002) illustrated the use of such representation. They showed that the ratio of the vegetative storage protein 1 (VSP1) encoding transcript to the VSP2 transcript increased when *A. thaliana* was wounded by biological agents, such as worms or moths, but not by mechanical wounding. However, in response to herbivore damage, both VSP1 and VSP2 transcript levels increased, while the ratio did not vary.

Yokoyama and Nishitani (2001) reported the expression analysis of 33 members of the xyloglucan endotransglucosylase/hydrolase (XTH) gene family in different organs of *A. thaliana*, as well as in response to four phytohormones. As mentioned in the previous section, the expression level of the 33 XTH genes varied a lot from one gene to another. Therefore, the expression profile of each gene was treated independently from the other genes, and the expression level of one given gene in root for example, was expressed as a percentage of the sum of the transcript level of the same gene in all the organs tested. This representation was necessary to illustrate the response of each gene to the tested treatments, regardless of their absolute level of transcripts. A similar strategy was adopted by Mladek *et al.* (2003) for the expression analysis of the *ARIADNE* gene in the different organs of *A. thaliana*. Nevertheless, it is worth noting that this type of data representation is biologically wrong if not all the plant organs were analysed, whatever their relevance to the issue of the study.

### 3.6.3-Compared to an endogenous gene

In numerous studies, the use of an endogenous gene to normalize the cDNA amount present in each sample is required. Amplification of this endogenous gene must be performed in experimental conditions identical to those of the target genes, in order to avoid technical errors. Therefore, multiplex PCR, where the control gene and the target gene are amplified in the same tube and at the same time, are strongly recommended for such studies.

#### - Constitutive genes

Normalization of transcript levels with a constitutively expressed gene allows to minimize the result of technical defects occurring during the reverse-transcription step. Different types of constitutive gene have been chosen for studies of plant gene expression. After cDNA synthesis with random oligonucleotides, numerous authors used the amplification signal corresponding to *A. thaliana* 18S (Itoh *et al.*, 2001; Klok *et al.*, 2002; Teramoto *et al.*, 2002; Hammond *et al.*, 2003; Liu *et al.*, 2003; Ogawa *et al.*, 2003; Shimada *et al.*, 2003; Stasolla *et al.*, 2003) or 28S (Choi *et al.*, 2002) ribosomal RNA. Alternatively, ribosomal proteins may

be used (Goda *et al.*, 2002). The ribosomes are very abundant in the cell, and using ribosomal RNA to normalize the transcript level of a weakly expressed gene may increase the deviation of the ratio. Moreover, as rRNAs and mRNAs are transcribed by different RNA-polymerases, their levels may not be paralleled, especially in some mutant backgrounds or some physiological stressing conditions (Bustin, 2002).

Instead, it appears more appropriate to use a constitutive gene whose level of expression is close to that of the gene of interest, such as the cytoskeleton-protein actin genes (Zheng *et al.*, 2002; Chae *et al.*, 2003; McGinnis *et al.*, 2003 and other), which, for example, showed absolute transcript levels twice higher than those of the GSK3/Shaggy kinases genes transcripts (Charrier *et al.*, 2002), or even the  $\beta$ -tubulin9 gene, which is expressed at a level twice to 100 times higher than the ARIADNE genes of *A. thaliana* (Mladek *et al.*, 2003). Other genes, known to be involved in house-keeping functions, may be chosen, such as the protein synthesis initiation factor eIF4A1 (Bonaventure *et al.*, 2003) or the protein synthesis elongation factor EF1- $\alpha$  (Collakova and DellaPenna, 2003; Foucher *et al.*, 2003), or in turn ubiquitin 2 (Itoh *et al.*, 2001) or ubiquitin 10 (Harari-Steinberg *et al.*, 2001; Wang *et al.*, 2003). Alternatively, Orsel *et al.* (2002) used the phosphate transporter gene *AtAPT1* to normalize the transcript level of the members of the NTR family involved in nitrate transport. Similarly, Bovy *et al.* (2002) used the tomato cyclophylin gene shown by Northern to be constitutively expressed in turning and red tomato fruit.

Jacobs *et al.* (2003) illustrated the simultaneous use of three constitutive genes to accurately normalize the expression of the glucan synthase-like genes in RNAi-transformed *A. thaliana* plants, as recommended by Vandesompele *et al.* (2002).

However, transcript level of so-called constitutive genes may vary under some growing conditions. For example, the level of *ACTIN2* transcript was reduced to approximately 50% when *A. thaliana* plants underwent a heat stress from root temperature to 37°C, while no variation in the same transcript level was observed at 44°C (Volkov *et al.*, 2003). Furthermore, constitutive genes may not adopt the same transcriptional behaviour in relation to the plant organ, or in response to growth changes. Indeed, the ribosomal protein L23 transcript level was reduced at 44°C, while that of *ACTIN2* did not vary compared to room temperature (Volkov *et al.*, 2003). Therefore, this is necessary to check whether the expression of the chosen constitutive gene is modified in response to the treatments tested on the target gene expression, as performed by Harari-Steinberg *et al.* (2001) with the ubiquitin10 gene in dark- or light-grown *A. thaliana* seedlings.

#### - Single copy gene

Studies intending to measure the number of copy of a transgene by real-time PCR need to normalize the level of genomic DNA. It can be performed with any endogenous gene, provided it is present as a single copy gene in the genome of the host plant. German *et al.* (2003) applied this strategy to determine the level of zygosity of transgenes in tomato by duplex PCR using the TaqMan technology. Using the same strategy, Schmidt and Parrott (2001) determined the number of copy of the cry1Ac transgene in soybean.

Measuring the level of transgenic or pathogene DNA contamination within cultivated plants also requires the use of an endogenous control. Hernandez *et al.* (2001) proposed the use of the acetylcarboxylase gene, which presents a high conservation level in 20 different varieties of *Brassica napus*, thereby avoiding optimisation of the protocol when changing rapeseed variety. Likewise, the cytochrome oxidase encoding gene has been used as an internal control for measurement of potato pathogen *Ralstonia* strains (Weller *et al.*, 2000).

Interestingly, in their study of the expression of the alternative oxidase genes during mango fruit ripening, Considine *et al.* (2001) validated both the quality and the quantity of their

cDNA preparations by testing a gene whose expression profile in the same conditions was known from previous experiments. This method was less accurate than using a constitutive gene. Nevertheless, it allowed to validate the ripening kinetics.

### 3.6.4-Compared to a sample of the same study

Some studies compared the expression level of one sample to that of another one. This is particularly relevant when the latter corresponds to a reference condition, such as the initial start point of a kinetic study (Charrier *et al.*, 2002; Brown *et al.*, 2003; Halterman *et al.*, 2003; Schenk *et al.*, 2003), or the control condition (Kursteiner *et al.*, 2003). Nevertheless, in order to ease the interpretation of the results, one sample can be arbitrary chosen to normalize the other ones. When studying expression profile of a gene in different organs, the value obtained for one of these organs can be chosen as the reference. For example, in the study of the expression level of the pyruvate decarboxylase genes in different organs of *A. thaliana*, the seedlings were arbitrary chosen as the reference organ, and levelled to 100 relative units (Kursteiner *et al.*, 2003). The reference value may correspond to the highest (Considine *et al.*, 2001) or one of the lowest values of the study (Jakab *et al.*, 2003). This method is particularly convenient when comparing expression patterns of several genes expressed with different levels and profiles (Shimada *et al.*, 2003).

## 4.0-APPLICATIONS

Since the beginning of the years 2000, real-time PCR has found a burgeoning number of applications in plant biology. A large variety of plants have been tested for either detection of foreign DNAs, or measurement of gene expression level. The range of plants used as biological material in real-time PCR experiments spans from green algae (*Chlamydomonas*, Teramoto *et al.*, 2002; Lake and Willows, 2003), bryophyte (the moss *Physcomitrella patens*, Richard O. and Charrier B., unpublished data), gymnosperm (*Pinus*, Anterola *et al.*, 2002; Stasolla *et al.*, 2003), angiosperm dicotyledonous plants used either for academic research, such as *Arabidopsis thaliana* (numerous reports), tobacco plants (Volkov *et al.*, 2003) and cell line (BY2, Dambrauskas *et al.*, 2003) and *Medicago truncatula* (Nolan *et al.*, 2003), or for applied purposes, such as tomato (Balbi and Lomax, 2003; Chang *et al.*, 2003; Liu *et al.*, 2003), potato (Svensson *et al.*, 2002), soybean (Maguire *et al.*, 2002), cacao (Jones *et al.*, 2002), dahlia (Suzuki *et al.*, 2002), peanut (Schmidt and Parrott, 2001), pea (Skaf *et al.*, 2000), sunflower (Thomas *et al.*, 2003), rapeseed (Hernandez *et al.*, 2001), cucumber (Ullanat and Jayabaskaran, 2002), mango (Considine *et al.*, 2001) and strawberry (Blanco Portales *et al.*, 2002). Many economically relevant monocotyledonous plants have similarly been used for real-time PCR experiments, such as rice (Banuelos *et al.*, 2002; Shen *et al.*, 2003), maize (Hahnen *et al.*, 2003; Philippar *et al.*, 2003), wheat (Sandberg *et al.*, 2003), barley (Choi *et al.*, 2002; Faccioli *et al.*, 2002; Halterman *et al.*, 2003) and bermuda grass (Faccioli *et al.*, 2002).

### 4.1- Detecting and quantifying the presence of foreign plant material.

#### 4.1.1-Food safety and agricultural concerns: detecting contamination with foreign DNA

As already stated at the beginning of this review, real-time PCR has proved to be a particularly useful technique for assessing the level of sample contamination by either GMPs

or pathogens. For such applications, the main limitation resides in the incapability of this technique to detect an organism whose sequence is unknown, since primer design requires DNA sequence information. Moreover, it is necessary to develop a new assay for each line of GMP (Kuribara *et al.*, 2002) or each microorganism.

- DNA from another wild-type plant

In some food quality control processes, checking for the absence of some undesired plant material is a necessary step before launching the product on the market. For example, the absence of gluten is of great importance for baby foods, and the presence of cereal genes will be searched by real-time PCR (Sandberg *et al.*, 2003). It may be as important for the agribusiness to reliably and quantitatively detect food ingredients for diagnosing adulteration. For example, spanish, italian and french regulations require that the maximum level of common wheat (*Triticum aestivum*) in durum wheat pasta be 3%, and real-time quantitative PCR was shown to be an ideal tool for assessing common wheat contamination in pasta and semolina (Alary *et al.*, 2002; Terzi *et al.*, 2003). In such cases, a major advantage of PCR-based quantification compared to ELISA tests is that DNA is more stable than proteins, and is therefore still present in highly processed products.

-DNA from genetically modified plants.

Culture of genetically modified plants (GMPs) raises concerns about contamination of the surrounding cultures and the storage infrastructure (Nap *et al.*, 2003). Therefore, consumer associations expressed their wish to see the presence of genetically modified product being checked and quantified before wild type species similar to those genetically modified were placed on the market. They also requested the establishment of a clear labelling on products derived from GMPs. Traditional PCR techniques were being widely used in this field, but at present, the need of truly quantitative techniques allowing to assess the level of contamination of food products is particularly critical since the European Parliament recently enforced a threshold of 0.9% for food products to be labelled as “produced from a genetically modified organism” (Nap *et al.*, 2003, EC Regulation N° 1829/2003).

Search for transgenic contaminations by real-time PCR has been performed on rapeseed (Zeitler *et al.*, 2002), but most of the applications aimed at detecting RoundUp Ready ® soybean, for example in soya flour (Berdal and Holst Jensen, 2001; Hird *et al.*, 2003) or transgenic maize (Vaitilingom *et al.*, 1999; Brodmann *et al.*, 2002). A common protocol can be applied for most of the maize transgenic lines, which targets the PCR oligonucleotides on common DNA regions, such as the *CaMV* 35S promoter (Hohne *et al.*, 2002). Alternatively, event-specific protocols may be designed for unique lines, as for the Starlink (Windels *et al.*, 2003), and the Bt11 (Ronning *et al.*, 2003) transgenic maize plants. In order to make the detection specific for only one given and identifiable event, scientists cloned the borders separating the transgene from the rest of the host genome, and used them as specific markers of the given event. So far, this strategy was applied for both maize (Hernandez *et al.*, 2003b) and soybean (Taverniers *et al.*, 2001) events.

Finally, the refinement of commercially available GMPs, by overexpressing endogenous genes or making genes silenced through antisense or RNAi strategies, will also make genetic modifications more difficult to detect, resulting in a need for techniques upgrading. For example, commercial PCR-based detection assays relying on the amplification of the *nptII* resistance gene or the *CaMV*35S promoter, are not relevant to detect GMPs with a deleted selection marker, or transgenes driven by tissue-specific promoter. Finally, naturally occurring events will have to be taken into account. For example, Weller *et al.* (2002)

reported the detection of numerous *Agrobacterium* strains in *Brassica napus* seeds, all of which devoid of the Ti plasmid. This reveals that PCR-based assays aimed at screening GMPs, and using primers designed on chromosomal sequences putatively common to a wide range of *Agrobacterium* strains, might lead to false positive results.

More details about GMPs screening are reported in the reviews of Auer (2003) and Ahmed (2002).

#### 4.1.2-Estimating transgene copy number in a plant genome

##### -Multiple insertions at different loci

Song et al. (2002) showed that real-time PCR could be used to estimate the number of transgenic loci present in the host genome. Quantification was achieved by adding a known amount (corresponding to a known number of copies) of plasmid recombinant for the transgene to the genomic DNA of the host plant. Alternatively, the standard might be a known unique sequence of the host plant genome. After completion of the PCR experiment, the signals between the standard and the samples are compared, and the number of copy of transgene integrated within the host genome was calculated, and found to be correlated with Southern blot results. However, this technique did not allow the authors to truly quantify the insertion number, because this application of quantitative PCR requires to discriminate reliably two fold changes in DNA amounts, which is best achieved using multiplex PCR (German *et al.*, 2003). Indeed, using a known unique sequence of the host genome as a standard in a duplex PCR assay, allowed to determine the transgene copy number in maize (Ingham *et al.*, 2001).

##### -State of zygosity of a transgene locus

In case of insertion at a single locus, the state of zygosity of a transgenic plant can be similarly determined by real-time PCR. The signal obtained after amplification of the transgene is compared to that of an endogenous gene present at the homozygous state. Using this technique, Schmidt and Parrott (2001) and German *et al.* (2003) were able to successfully determine the level of zygosity of transgenes in tomato, soybean and peanut.

### **4.2-Detecting and quantifying plant parasites and symbiotes.**

#### 4.2.1-Health monitoring of cultivated species

Another useful application of real-time quantitative PCR is the assessment of contamination of crops, seeds or food by plant pathogens or harmful microorganisms. Although still regarded as costly and skill demanding, real-time quantitative PCR is currently one of the best alternatives for the diagnostic of diseases in field-studies. For example, Winton *et al.* (2003) compared four methods to quantify the colonization of Douglas-fir foliage by the fungus *Phaeocryptopus gaeumannii*. Although all methods gave significantly correlated results, they concluded that real-time PCR was more rapid and less cumbersome than fructification counting, and also allowed to discriminate the severity of infection before the fungus set fruit. It was also more specific than ergosterol quantification, since ergosterol is a lipid produced by higher fungi and is therefore not specific for a single species. Finally, real-time PCR was much more sensitive than dot-blotting DNA with a specific probe. Schaad *et al.* (2002) also insisted on the fact that real-time PCR allowed to diagnose Pierce's disease of grape long before symptom appearance, which offers better perspectives for controlling the disease in the

field. In one of the very first reports of the use of TaqMan technology in plant virus detection, Schoen *et al.* (1996) presented an assay, which reduced the diagnosis time for potato infection by Potato Leafroll Virus from five weeks to one day, compared to the previously used method. This example also illustrates that nucleotide based assays are a serious alternative when ELISA-based detection methods are difficult to apply, due to some particularities of the model. Likewise, because viroids are devoid of capsid, antibodies cannot be raised against this type of pathogenic agent. Moreover, a PCR-based detection system is easier to develop than a protein detection assay, since it does not require raising specific antibodies against the microorganism. Besides, compared to classical PCR techniques, real-time quantitative PCR does not require post-PCR handling, which saves time and also decreases the risk of sample contamination.

Thanks to its sensitivity, real-time PCR is well suited for rapid and accurate quantification of microorganisms in crude extracts, like soil samples (Filion *et al.*, 2003a and references therein). Filion *et al.* (2003b) also reported that in such cases, real-time quantitative PCR brings more robust results than cfu counting on semi-selective medium, probably because of a more homogenous extraction yield. Another way to achieve an even higher sensitivity is to design primers targeted at multicopy sequences like rRNA-encoding genes. Indeed, many fungal detection assays are based on the amplification of internal transcribed spacer regions (ITS), which are both specific for species and highly repeated in the genome (e.g. Cullen *et al.*, 2001; Schena *et al.*, 2002; Atkins *et al.*, 2003).

In addition, the high specificity of this technique allows critical identification of morphologically similar microorganisms as well as quantification of biotrophic pathogens and of obligate symbiots that are impossible to grow *in vitro*. It should be noted however that real-time quantitative PCR is complementary to other identification methods, since it can only be performed on previously identified organisms (Heuser and Zimmer, 2002). Besides, due to the very small size of the samples used, a preliminary enrichment of microorganisms on semi selective medium might be useful when a very high sensitivity is required. During this process, naturally occurring PCR inhibitors are also removed, thus limiting the risk of false negative results. Such methods, combining growth on agar plates or in liquid media and real-time PCR quantification, have been developed for a number of bacterial pathogens such as *Agrobacterium* strains (Weller and Stead, 2002). In this case, the authors reported that a sensitivity of 10 cfu/mL could be achieved after a 48h enrichment of the samples. Other examples of the use of this technology are reported in Schaad *et al.* (2003).

Discrimination between closely related strains is also possible, based on known nucleotide polymorphism. For instance, food contamination with fungal toxins is becoming a major concern in food safety, and it is known that toxin content is not correlated with fungus abundance, but rather dependent on the toxigenic properties of the strains. Schnerr *et al.* (2000) introduced a very rapid method for quantifying the *tri5* gene from *Fusarium*, which encodes a key enzyme of trichotecene biosynthesis, therefore assessing the risk of contamination of each sample. Such a protocol may be of great interest since the EU is about to introduce a regulation concerning the acceptable level of cereal contamination by these toxins.

Finally, the specificity guaranteed by real-time PCR allows to monitor several microorganisms at the same time on the same sample, provided different sets of primers are used. This is particularly convenient to address the competitive or synergistic relationship between microorganisms living in association with the same plant, which is an important issue of ecology (Eun *et al.*, 2000; Filion *et al.*, 2003b).

Because real-time quantitative PCR can be transposed to virtually every pathosystem, the number and the diversity of published applications in this domain has been growing exponentially the last few years, as illustrated by the references below. While the first application of real-time quantitative PCR in phytopathology was published in 1996 by Schoen *et al.*, which reported the detection of Potato Leafroll Virus (PLRV) in potato tubers, many assays have been developed since then. For example, the detection of phytoviruses such as the Apple Chlorotic Leaf Spot Virus (ACLSV) (Salmon *et al.*, 2002), the Tomato Spotted Wilt Virus (TSWV) on different crops (Roberts *et al.*, 2000), the Dahlia Mosaic Virus (DMV) on *Dahlia pinnata* (Nicolaisen, 2003), and of the episomal Banana Streak Virus (BSV) integrated in banana and plantains genomes (Delanoy *et al.*, 2003), was eased by the use of this technology. Some studies have focused on a particular aspect of viral biology, like the virus infectivity (Skaf *et al.*, 2000), or its transport by insect vectors (Goncalves *et al.*, 2002). Multiplex assays have also been developed to discriminate between virus causing similar symptoms or coexisting in a single plant, like the Potato Mop Top Virus (PMTV) and the Tobacco Rattle Virus on potato (Mumford *et al.*, 2000), or the Cymbidium Mosaic Potexvirus (CymMV) and the Odontoglossum Ringspot Tobamovirus (ORSV) on the orchid *Oncidium* (Eun *et al.*, 2000).

Regarding fungi, real-time quantitative PCR has already been applied to numerous interactions such as: *Phytophthora citricola* on oak and beech (Böhm *et al.*, 1999), *Botrytis cinerea* on grapevine leaves and berries (Bezier *et al.*, 2002), *Fusarium solani* and *Gomus intraradices* on common bean (Filion *et al.*, 2003b), *Phytophthora infestans*, *Rhizoctonia solani*, *Helminthosporium solani* and *Colletotrichum coccodes* (Böhm *et al.*, 1999; Cullen *et al.*, 2001, 2002; Lees *et al.*, 2002) on potato tubers, the late blight causing agent *Phytophthora infestans* in potato and tomato plants (Avrova *et al.*, 2003), and the rust fungus *Phakopsora pachyrhizi* in soybean (Frederick *et al.*, 2002). In monocotyledonous hosts, the possible use of this technique to detect crop-threatening fungi has similarly been investigated. Assays were developed for *Magnaporthe grisea* in rice (Qi and Yang, 2002), for *Pyrenophora* on barley seeds (Bates *et al.*, 2001) and for a numerous wheat pathogens (Fraaije *et al.*, 2001, 2002). An assay was similarly developed on *Plectospharella cucumerina*, a nematophagous fungus that could potentially be used as a biocontrol agent of two devastating nematode species (Atkins *et al.*, 2003).

Finally, the bacterium *Clavibacter michiganensis* was quantified in potato stems and tubers (Mills and Russell, 2003). Likewise, an assay was developed for the brown rot agent *Ralstonia solanacearum* in potato tubers (Weller *et al.*, 2000). It provides a fast and reliable alternative to the procedures routinely used for seed control since the sale of *Ralstonia*-infected seed potatoes is prohibited in the EU (Council directive 2002/56/EC on the marketing of seed potatoes). *Xylella fastidiosa* was detected in citrus trees (Oliveira *et al.*, 2002), in grape stems (Schaad *et al.*, 2002), and in orange fruits and seeds (Li *et al.*, 2003), and *Polymyxa* in sugar beet (Kingsnorth *et al.*, 2003).

The presence of different *Agrobacterium* strains and closely related bacteria was detected by the amplification of the Ri plasmid in known root mat previously studied with more conventional protocols requiring complex DNA extraction (Weller and Stead, 2002).

A detailed review on the applications and prospects of real-time quantitative PCR among other available techniques in plant fungal disease diagnostic is provided by McCartney *et al.* (2003). Schaad and Frederick (2002) and Schaad *et al.* (2003) also gives an exhaustive survey of the current pathogen detection assays based on real-time quantitative PCR .

#### 4.2.2-Studies of plant response to pathogen infection

Apart from being used for diagnosis in field studies, real-time quantitative PCR is also becoming the reference technique for following the progress of infection in more fundamental studies. This allows to monitor the sensitivity of mutants or transgenic plants to a special pathogen or conversely to address the infectivity of a pathogen. For example, we have set up a real-time quantitative PCR test for monitoring the growth of two pathogenic fungi *Alternaria brassicicola* and *Botrytis cinerea* on *Arabidopsis thaliana*, and control experiments have shown the protocol to be fast, sensitive, reproducible and robust (C. Gachon and P. Saindrenan, unpublished results). Comparable tests are being established on a number of models, and they are far more convenient than classical resistance tests based on conidiophores or colony counting, or visual appreciation of symptoms (Ito *et al.*, 1998; van Wees *et al.*, 2003). A very appreciable feature is that samples might be frozen and analysed later, whereas symptoms have to be scored immediately during the time course.

Considering the minute quantities required for performing an expression assay (10 to 20 ng compared to about 10 µg for a Northern blot), it is also possible to monitor the expression of numerous genes after a single RNA extraction, even if the amount of starting material is limited. In the case of plant-microorganism interactions, it has been used for studying the coordinate expression of both microbial and plant genes in order to better characterize the interaction (Guescini *et al.*, 2003).

#### **4.3-Studying the expression profile of multigene families**

Gene families are a feature of plant genomes. Indeed, animal genomes present very rare gene families. The analysis of the genome of *A. thaliana* revealed that 66% of the genes belong to gene families, with half of them containing at least five members (AGI, 2000). In addition, if members of gene families originated from recent duplications, their paralogues may present a very high level of identity at the nucleotide level. Therefore, in order to have an accurate view of the function of the family, several studies described its overall expression profile by real-time RT-PCR. In the past, plant scientists had recourse to RT-PCR methods rather than Northern blot, because first, genes expressed at very low level are very difficult to detect by Northern blot (Jakab *et al.*, 2003), and secondly, members of gene families are usually very conserved at the sequence level, making cross-hybridisations likely to occur during Northern blot procedures (Montrichard *et al.*, 2003). In this section, we propose a review of the studies reporting gene family expression analyses performed by real-time RT-PCR. This review is not exhaustive, as it aims at illustrating the capability of real-time RT-PCR to meet the need for specificity, sensitivity, rapidity and overall accuracy, making possible comparison between organisms and allowing speculations on molecular and transcriptional evolution of the different gene family members.

First of all, real-time RT-PCR may be a required tool because of the limitations encountered with other more commonly used techniques, such as semi-quantitative RT-PCR. For example, the study of the whole *A. thaliana* family of NTR2 nitrate transporters is an illustration of the combination of the use of semi-quantitative RT-PCR and real-time RT-PCR for analysing the dynamics of a gene family in plants. The transcript level of the seven members of the NTR2 gene family were tested by semi-quantitative PCR in both seven different plant organs and on two different supplies of nitrate (Orsel *et al.*, 2002). Because the quantification performed by semi-quantitative PCR was directly dependent on the hybridisation efficiency of the probe to the PCR product, the authors subsequently used real-time PCR, in order to confirm the expression profile observed in leaves and roots of adult plants grown in limiting nitrate.



Hence, the main interest of real-time RT-PCR for the study of gene family expression remains its high specificity. The analysis of the expression profile of the 33 members of the xyloglucan transglucosylase hydrolase (XTH) gene family required a high specificity in the detection of the different XTH members (Yokoyama and Nishitani, 2001). The expression level of all the members of this *A. thaliana* XTH gene family was examined in both different plant organs and in response to several hormonal treatments. Some members of this gene family display 90% identical nucleotide sequences, requiring the use of RT-PCR technique rather than Northern blot. In addition, cDNA synthesis was performed by an oligonucleotide specific for each of the 33 members. The absolute level of transcripts varied considerably from gene to gene. Indeed, while 12 members displayed at least 2000 copies of transcript per nanogram of total RNA in root, leaf, stem, flower bud and green silique, at least seven members were expressed at the background level. A relative analysis of the data showed that the expression of 10 XTH members were root-specific, while five were silique-specific, two stem-specific and two flower-specific, the other genes displaying a quasi-constitutive expression pattern. In response to hormones, XTH genes displayed different expression profile, some being induced by both auxin and brassinolide or gibberellic acid, other repressed by auxin or abscisic acid. The analysis of the expression profile was compared with the sequence similarity level between the XTH genes. It revealed that genes originating from recent duplications exhibited similar expression profiles, but not similar expression levels (Yokoyama and Nishitani, 2001).

The GSK3/Shaggy kinases are serine/threonine kinases, coded by 10 genes in *A. thaliana*, and named *AtSK*. Their level of expression is rather low, and difficult to study by Northern blot, yet possible, as reported by Dornelas *et al.* (1999). However, some of these genes display a very high level of sequence identity at the nucleotide level, and finding probes specific for each of the 10 genes requires to design them in the 5' or 3' UTR regions. In order to get round these two difficulties, real-time RT-PCR has been used to study the expression profile of the whole gene family, both in different organs of the plant, and in response to a series of abiotic stress. Relative analyses showed that the 10 members were all expressed at similar levels whatever the plant organs. They displayed no organ specificity, except *AtSK31*, which is flower specific. By contrast, in response to a series of abiotic stress, two genes displayed a significant induction to hyperosmotic stress, and one was induced by darkness (Charrier *et al.*, 2002). Quantitative analyses, using *A. thaliana* genomic DNA as a quantitation standard, revealed that the *AtSK* genes were expressed at moderate absolute level, *i.e.* twice lower than the *ACTIN2* gene (Charrier *et al.*, 2002).

Interestingly, the homologous gene family has been studied in the moss *Physcomitrella patens* (Richard O. and Charrier B., unpublished data). Despite that the sequence of the whole genome of *P. patens* is not currently known, we showed that the GSK3/Shaggy gene family is likely to be smaller in *P. patens* than in *A. thaliana*, with putatively only five members (called *PpSK*). The genomic analysis of these five genes revealed that they are all highly conserved, both at the amino acid level, and at the genomic organisation level, with the position and the size of the introns conserved between the members. Studying the functions of the GSK3/Shaggy kinase genes in *P. patens* will provide information about the function of the GSK3/Shaggy genes present in the ancestor of both bryophytes and angiosperms. Because, on the one hand, all the *PpSK* genes may not be identified, and on the other hand, they present a very high level of structure and sequence similarity, specific detection of the five *PpSK* genes identified so far must be guaranteed. Therefore, their expression analysis was performed by real-time RT-PCR. The data showed that the *PpSK* genes were induced by hyperosmotic stress, as some of the *AtSK* genes, and their absolute transcript level was similar to that of the *AtSK* (Charrier *et al.*, 2002).

This comparison was made possible only because of the accuracy of the data due to the technical features of real-time RT-PCR. It allowed to reveal that the GSK3/Shaggy have conserved common functions from bryophytes to angiosperms.

Another quality of real-time RT-PCR for the study of gene family expression profiles is its sensitivity. The aminocyclopropane-1-carboxylic acid synthase (ACS) enzyme is a key regulatory enzyme of ethylene biosynthesis, and hence is involved in fruit maturation in tomato (*Lycopersicon esculentum*). The expression pattern of the 8 known members of the ACS gene family was investigated by real-time RT-PCR in tomato plants. The level of expression varied significantly between members of this gene family (roughly 10-fold) (Balbi and Lomax, 2003), and members, previously considered as not expressed from Northern blot or ribonuclease protection assays experiments, displayed this time an extremely low transcript level, yet made detectable by real-time RT-PCR (Balbi and Lomax, 2003).

The pyruvate decarboxylase enzyme (PDC) is responsible for the first step in the ethanolic fermentation. In *A. thaliana*, the family is composed of four genes, whose expression in different organs and in response to anoxia and other environmental stresses has been studied by real-time RT-PCR (Kursteiner *et al.*, 2003). The transcript level of these four genes had previously been shown to be low, justifying the use of RT-PCR techniques. Among the four *PDC* genes, *PDC1* was induced at a high absolute transcript level in response to anoxia only. Indeed, the other abiotic stresses, such as osmotic, salin and oxidative stresses, cold and wounding, as well as the treatment to abscisic acid, did not modify the transcript level of *PDC1*. The three other genes were insensitive to all the treatments applied. In the same study was reported the expression profile of the aldehyde dehydrogenase (ALDH) gene family. ALDH enzyme catalyses the oxidation of aldehydes, leading to its detoxification by production of less reactive forms, and its putative role in anoxia resistance has been tested by studying the expression profile of the gene family in fermentation condition. Because the transcripts of the three *A. thaliana* *ALDH* genes are present in a high copy number, the expression study was performed by Northern blot analysis instead of real-time. This study illustrates the limitation of real-time PCR. Indeed, its cost makes it a dispensable tool for studying genes strongly expressed and sufficiently different from their homologues at the sequence level.

In addition to its high sensitivity, the accuracy of real-time PCR along several orders of magnitude, was exploited for the study of the expression profile of the 9-cis epoxycarotenoid dioxygenase. This enzyme cleaves 9-cis xanthophylls to xanthoxin, which is a precursor of abscisic acid. In *A. thaliana*, the gene family comprises five members (*AtNCED*). Real-time RT-PCR (TaqMan), coupled to *in situ* detection of transcriptional activity by promoter::*uidA* fusion, allowed to describe a detailed expression pattern of these genes in different organs and tissues of the plant (Tan *et al.*, 2003). Depending on the organ tested, the absolute level of transcripts of the *AtNCED* genes ranges from 20 copies per nanogram of total RNA (*AtNCED9* in roots, which is close to the background level), to 1,200 copies for *AtNCED3* in secondary stems, revealing an important discrepancy between genes with regard to their absolute transcript level. The latter was not correlated to the distance relationship between *AtNCED* genes. For example, *AtNCED 3* and *9* are very close to each other, likely to originate from a recent duplication, while *AtNCED3* is expressed at the highest level and *AtNCED9* among the lowest in roots and stems. Likewise, *AtNCED2* and *AtNCED5* display different organ specificity of expression, while they originated from a very close duplication, revealing a lack of correlation between the expression profile and the relationship between genes. The combination of the data led to the conclusion that the biosynthesis of ABA takes place in different organs, and is developmentally regulated.

The ARIADNE gene family is another example of family requiring the sensitivity of real-time RT-PCR added to its capacity for quantifying expression level with large discrepancies. It was identified from its homology with *Drosophila* and mouse genes involved in the ubiquitin-mediated protein degradation pathway, and thought to be E3-ligases. Sixteen genes compose this family in *A. thaliana*, and their expression was studied by real-time RT-PCR (Mladek *et al.*, 2003). Among this family, the expression of seven genes is undetectable or close to the background level. All the other genes are expressed with very different absolute transcript levels, spanning along a range of 25-fold variation. This absolute transcript level remained very low, with less than 150 copies per nanogram of total RNA. While the expression profile was quasi-constitutive for most of the *AtARI* genes, two genes displayed a specific expression pattern. *AtARI12* was root specific, and *AtARI16* was silique specific. No correlation could be done between the position of the gene within a phylogenetic tree, and both the absolute transcript level and the transcript profile (Mladek *et al.*, 2003).

Other studies illustrating the use of real-time RT-PCR for describing the expression profile of gene family members in plants or in algae were recently reported (Panchuk *et al.*, 2002; Reintanz *et al.*, 2002; Teramoto *et al.*, 2002; Shimada *et al.*, 2003).

In conclusion, real-time RT-PCR is often an unavoidable tool required to perform an accurate study of the expression profile of related gene family members.

#### **4.4-Confirmation of large-scale expression studies**

The reliability of micro-array experiments may sometimes be discussed, especially for cDNA-based or PCR fragment-based chips. As plants display a high number of multigene families, cross-hybridisation between cDNA representatives of gene family members may lead to false interpretation. Because of its high accuracy, real-time RT-PCR is now extensively used to verify and confirm the induction ratio calculated from micro-array data. In most cases, it confirmed the gene induction or repression previously detected by micro-array experiments, with a linear correlation up to 5 orders of magnitude (Maguire *et al.*, 2002). However, several other studies reported that the n-fold variation measured by real-time PCR was generally lower (up to 10 times) than that measured by micro-array (Klok *et al.*, 2002; Schenk *et al.*, 2003). Interestingly, Wang *et al.* (2003) illustrated that real-time data can also display induction ratios higher than micro-arrays, with still a good correlation between the two techniques.

Even with the Affimetrix chips, which are oligonucleotide-based and therefore much more specific than PCR fragment based micro-array slides, discrepancies between micro-array data and real-time PCR data have been observed. On the one hand, Rider *et al.* (2003) picked randomly 12 candidates among a pool of 185 genes, previously identified by Affimetrix micro-array experiments as being up-regulated in the seeds of the *Arabidopsis* mutant *pickel* compared to wild type seeds. They showed that approximately 80% of these genes displayed a confirmed up-regulation by real-time RT-PCR experiments. On the other hand, Hammond *et al.* (2003) noticed that not only the magnitude and the kinetics of the response of several genes to phosphate starvation differed between the two techniques, but for some of them, the response was also opposite. They explained this result by the fact that the two experiments were performed with two different biological samples and different oligonucleotides.

In all cases, real-time RT-PCR is considered as the most reliable technique, and may even be used to test, after comparison of the final data, different methods of analyses of micro-array data (Puthoff *et al.*, 2003). However, while micro-array experiments analyse up to thousands of genes in one step, real-time PCR is limited to only few tens of genes. Indeed, it requires the

design of oligonucleotides specific for each gene analysed, and because of the limited number of both fluorophores and light spectra detected by real-time PCR machines, this technique allows the detection, by multiplex PCR, of maximum five genes at a time. In routine however, a maximum of two genes are commonly analysed in the same tube. Therefore, only genes for which an interesting response was revealed by micro-array experiments, are prone to confirmation by real-time PCR.

While some of these studies are purely confirmatory (Ogawa *et al.*, 2003), other ones are using real-time RT-PCR to further analyse the expression pattern of the candidate genes, either along fine kinetics (Goda *et al.*, 2002; Goto and Naito, 2002), or in conditions for which few material is available, thereby taking advantage of the sensitivity of this technique. By hybridisation of cDNA micro-arrays with RNA of *Arabidopsis* seedlings infected with the incompatible fungus *Alternaria brassicicola*, Schenk *et al.* (2003) identified functional groups of genes involved in systemic acquired resistance. Twenty three genes, each representative of one of the functional groups, were chosen to perform a time-course study in response to *A. brassicicola* infection in local and distal leaf tissues by real-time RT-PCR (Schenk *et al.*, 2003).

Other types of large-scale studies were confirmed by real-time PCR. Chen *et al.* (2003) analysed the transcriptional response of *Arabidopsis* to genotoxic stress by screening a high density colony array (HDAC). Out of 27,000 cDNA clones present in these colonies, 142 were identified as being up- or down- regulated by genotoxic agents, out of which 42 were selected for real-time PCR analysis. The response to genotoxic stress was confirmed for 80% of these 42 clones. Likewise, *Arabidopsis* genes responding to the nitroaromatic compound 2,4,6-trinitrotoluene, were searched by serial analysis of gene expression (SAGE). Three genes, each representative of the three main classes of transcriptional responses previously identified by SAGE (induced, repressed, or unaffected levels), were tested by real-time RT-PCR. Comparisons of the data showed a good correlation between these two techniques (Ekman *et al.*, 2003).

#### **4.5- Molecular phenotype of mutant and transgenic plants**

##### **4.5.1-Measuring the effect of a mutation on its target gene**

Monitoring the transcriptional level of a mutated gene or of a transgene present in plant, is the first step towards the full analysis of their effect of plant biology. In these contexts, several studies reported the use of real-time RT-PCR to quantify the transcripts level of a transgene (Itoh *et al.*, 2001; Busch *et al.*, 2002; Chang *et al.*, 2003), or to demonstrate that a mutated gene was effectively transcriptionally inactivated (Bonaventure *et al.*, 2003). Jacobs *et al.* (2003) monitored the effect of three RNA interference constructs targeting glucan synthase-like genes in *A. thaliana*, and showed that the level of the target transcripts was significantly reduced in the transgenic plants. Likewise, the presence of a 40-60 bp inverted-repeat, specific for two genes of interest and introduced into plants *via* a viral vector, was shown to increase the efficiency of the so-called Virus-Induced Gene Silencing (VIGS) mechanism up to three times, compared to a vector with only a sense or an antisense specific cDNA sequence (Lacomme *et al.*, 2003).

In tomato, Liu *et al.* (2003) investigated the influence of the expression level of the quantitative trait locus fw2.2 on the tomato fruit mass. They constructed a gene dosage series by introducing into tomato plants 0 to 4 fw2.2 alleles by genetic crosses and transformation.

The increase in the fw2.2 transcript level of the generated plants was measured by real-time RT-PCR, and showed to be negatively correlated to the fruit mass.

#### 4.5.2-Assessing the impact on other gene expression

As determining the molecular phenotype of a mutant is usually performed without prejudice, micro-array experiments are often used for this type of studies. The *Arabidopsis* sulfate transporters SULTR1 and SULTR2 ensure the sulfate uptake activity at the root surface. Maruyama-Nakashita *et al.* (2003) characterised a novel mutant with a *SULTR2* allele disrupted by a T-DNA insertion. Micro-array experiments identified genes, whose expression level was modified as a result of the mutation, and this change in the expression level was then verified by real-time RT-PCR.

Real-time RT-PCR may also be used for lower scale studies, without requiring a previous micro-array experiments screening. Rider *et al.* (2003) investigated whether the expression pattern of genes known or suspected to be involved in embryogenesis in *A. thaliana* were co-regulated in the mutant background *pickle* (*pk1*), which displayed a disturbed seed germination process. Real-time RT-PCR experiments allowed the authors to classify these genes within two groups, those whose expression was regulated by PKL, and those which were insensitive to this mutation. Likewise, in tomato, the expression of Aux/IAA auxin-responsive genes, as well as members of the ACS gene family, a key regulatory enzyme of ethylene biosynthesis, was compared between wild type tomato plants and the mutant *diageotropica* displaying an alteration of fruit development (Balbi and Lomax, 2003). The results of real-time RT-PCR experiments suggested that the early stages of fruit development in tomato were controlled by auxin- and ethylene-dependent mechanisms.

Molecular effect of transgenes on endogenous gene expression may also be investigated. Tomato plants transformed with the *Agrobacterium rolD* gene displayed an increased tolerance to toxin compared to wild-type plants. Real-time RT-PCR allowed to show that the pathogenesis-related gene *PR1* was also up-regulated in response to salicylic acid, thereby confirming that *rolD*-transformed tomato plants displayed an increased competence for defence response (Bettini *et al.*, 2003). Likewise, Brown *et al.* (2003) aimed at investigating whether the overexpression of the *A. thaliana* jasmonate/ethylene responsive gene *AtERF2-1* had an effect on the expression level of pathogen-responsive genes containing a GCC-box within their promoters. Real-time RT-PCR showed that at least three of them were overexpressed in transgenic lines, and that the GCC motif was likely to interact with the ERF transcription factors.

Finally, Busch *et al.* (2002) used real-time RT-PCR to monitor the effect of antisense expression on the transcript level of genes involved in the same metabolic process. They showed that, while the expression of an antisense construct specific for the tobacco phytoene synthase *PSY1* encoding gene moderately reduced the steady-state level of *PSY1* transcripts, it reduced the level of the homologous gene *PSY2* to 25% of its initial level.

#### 4.6-Miscellaneous

Numerous studies are now using real-time RT-PCR to establish the relative expression profile of only one gene of interest, in different organs or in response to different treatments (Blanco Portales *et al.*, 2002; Choi *et al.*, 2002; Holmberg *et al.*, 2002; Suzuki *et al.*, 2002; Dambrauskas *et al.*, 2003; Nolan *et al.*, 2003; Shen *et al.*, 2003; Thomas *et al.*, 2003) or in turn in different genetic backgrounds (Foucher *et al.*, 2003). The sensitivity of real-time PCR is exploited particularly for those of these genes presenting a very low steady-state transcript

level. However, in laboratories with no budget limitation, real-time RT-PCR may also be used only because it provides reliable results in a short time.

Yet, a number of studies illustrated some original applications of real-time PCR in fundamental plant biology. For example, in maize, the editing process has been shown to affect the expression of 15 different plastid genes. Peeters and Hanson (2002) investigated the contribution of editing to the regulation of chloroplast gene expression, compared to the abundance of transcript itself. The relative transcript level of 10 genes was measured by real-time RT-PCR, and compared to the editing rate. This study allowed to demonstrate that the efficiency of the editing process of these genes was not dependent on their transcript abundance. In the domain of transcript level regulation by epigenetic mechanisms, Jones *et al.* (2001) monitored the level of transcript corresponding to a transgene, by quantifying the proportion of methylated DNA resulting from post-transcriptional gene silencing.

Alternatively, some interesting studies reported the expression analysis of several unique genes all involved in the same physiological process. In *A. thaliana*, the expression of eight genes all involved in the brassinosteroid (BR) biosynthesis, degradation or reception pathways was investigated by real-time RT-PCR. The results allowed to correlate the expression of the BR synthesizing enzymes with the presence of BR *in planta*, and to have a more integrative view of BR metabolism in the whole plant body (Shimada *et al.*, 2003). Likewise, Anterola *et al.* (2002) studied all known genes involved in the biosynthesis of lignin-lignan precursors in Pine tree, *i.e.* seven enzymes (PAL, C4H, 4CL, CAD...). The coordinated response of some of these enzymes to an increasing concentration of phenylalanine allowed to better understand the integrated regulation of this secondary metabolism biosynthesis pathway.

## 5.0- LIMITATION OF REAL-TIME PCR TECHNIQUES IN PLANT RESEARCH

Despite its numerous advantages, real-time PCR presents also some limitations. First, from a fundamental point of view, the amplification of fragments < 200 bp can mislead the interpretation of the result, especially for RNA expression analyses. Indeed, detection of only a minor part of the sequence may not reflect the whole sequence. For example, several transcripts with different sizes may be expressed from the same gene, due to alternative splicing. In case the splicing sites are known, and alternative splicing previously displayed by Northern blot, then real-time RT-PCR can monitor the alternative splicing event after appropriate oligonucleotides hybridising with the introns are designed, as nicely illustrated by Halterman *et al.* (2003). Otherwise, real-time PCR is unable to detect alternative splicing, and will lead to inaccurate interpretation of the data. Furthermore, partial transcript degradation may occur, which is especially relevant to gene silencing studies. In all those cases, real-time RT-PCR should not be the only tool used, and Northern blot analyses should complement the data. Montrichard *et al.* (2003) studied by real-time RT-PCR the absolute expression level of the NADPH/NADP-thioredoxin TRXh3 encoding gene in pea, and in parallel, displayed by Northern blot that the gene was likely to be coded by two transcripts with different size.

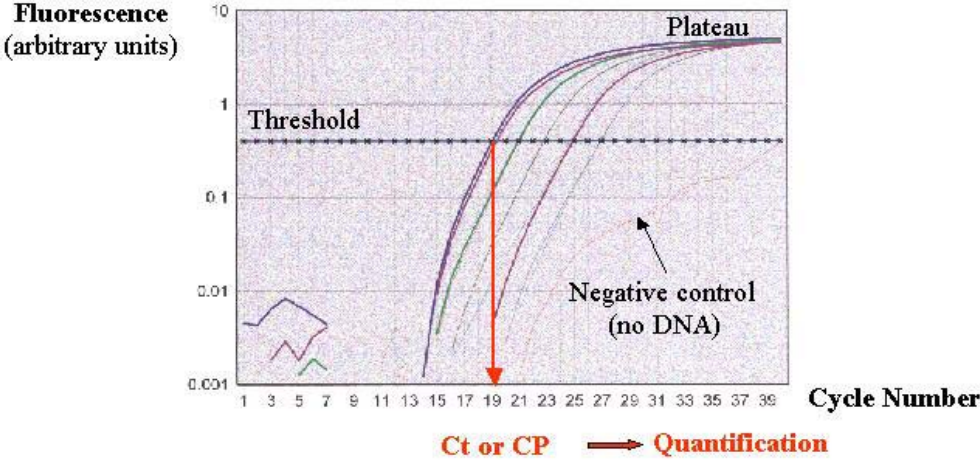
Secondly, from a more practical point of view, real-time RT-PCR analyses are usually performed at the tissue level only, because several micrograms of RNA are typically needed. Even if Philippar *et al.* (2003) used real-time RT-PCR to successfully compare the expression profile of a potassium channel encoding gene in epidermis, mesophyll and vascular tissues of the maize leaf, further studies using fusions of promoter to reporter genes may be necessary to complement the description of the expression profile (Berger *et al.*, 2002; Reintanz *et al.*, 2002).

Finally, compared to other more conventional techniques, real-time (RT)-PCR presents additional steps. DNA has to be extracted and submitted to a high number of PCR cycles, and reverse-transcription may bring additional biases between samples. Therefore, this technique requires a rather high number of repeats to be statistically reliable. The number of repeats varies from one worker to another one (Bustin, 2002), and depends on the study performed. The cost of the whole experiment largely depends on this number of repeats.

## 6.0- CONCLUSION

This review illustrates to what extent real-time PCR is useful to measure both DNA and RNA levels in plants, in both an absolute and a relative way. Classical PCR or RT-PCR used to fulfil these aims, and still do. Yet, real-time PCR became an as attractive as necessary tool, because it can ensure reliability while being extremely rapid, specific while ensuring a high sensitivity, making the combination of all these features necessary for more and more studies. However, comfort and peace of mind with regard to the results have a price to pay. Because of the cost of both the machine and the fluorophores, real-time PCR is now just emerging as a routine tool in plant laboratories.

**Typical PCR amplification curves**



**Figure 1: Illustration of real-time PCR amplification curves**  
Increase of fluorescence along PCR amplification is plotted against the number of PCR cycles. Curves, corresponding to templates with increasing concentration of DNA from the left to the right, describe an exponential amplification phase, followed by a plateau. A threshold of fluorescence is set in the exponential phase of the reaction, above the residual fluorescence signal of the negative control. The number of cycles necessary to cross this threshold line is defined as the Ct (threshold cycle) or CP (crossing point), and is used to calculate the initial template concentration (see text and references therein).



## Bibliography

- AGI. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**: 796-815.
- Ahmed, F.E. 2002. Detection of genetically modified organisms in foods. *Trends Biotechnol*, **20**: 215-223.
- Alary, R., A. Serin, M.P. Duviau, P. Joudrier and M.F. Gautier. 2002. Quantification of common wheat adulteration of durum wheat pasta using real-time quantitative polymerase chain reaction (PCR). *Cereal Chem*, **79**: 553-558.
- Anterola, A.M., J.H. Jeon, L.B. Davin and N.G. Lewis. 2002. Transcriptional control of monolignol biosynthesis in *Pinus taeda* - Factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *J Biol Chem*, **277**: 18272-18280.
- Arumuganathan, K. and E.D. Earle. 1991. Nuclear DNA content of some important plant species. *Plant Mol Biol Rep*, **9**: 208-218.
- Atkins, S.D., I.M. Clark, D. Sosnowska, P.R. Hirsch and B.R. Kerry. 2003. Detection and quantification of *Plectosphaerella cucumerina*, a potential biological control agent of potato cyst nematodes, by using conventional PCR, real-time PCR, selective media, and baiting. *Appl Environ Microbiol*, **69**: 4788-4793.
- Auer, C.A. 2003. Tracking genes from seed to supermarket: techniques and trends. *Trends Plant Sci*, **8**: 591-597.
- Avrova, A.O., E. Venter, P.R.J. Birch and S.C. Whisson. 2003. Profiling and quantifying differential gene transcription in *Phytophthora infestans* prior to and during the early stages of potato infection. *Fungal Genet Biol*, **40**: 4-14.
- Balbi, V. and T.L. Lomax. 2003. Regulation of early tomato fruit development by the *diageotropica* gene. *Plant Physiol*, **131**: 186-197.
- Banuelos, M.A., B. Garcíadeblas, B. Cubero and A. Rodríguez Navarro. 2002. Inventory and functional characterization of the HAK potassium transporters of rice. *Plant Physiol*, **130**: 784-795.
- Bates, J.A., E.J.A. Taylor, D.M. Kenyon and J.E. Thomas. 2001. The application of real-time PCR to the identification, detection and quantification of *Pyrenophora* species in barley seeds. *Mol Plant Pathol*, **2**: 49-57.
- Becker-Andre, M. and K. Hahlbrock. 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Res*, **17**: 9437-9446.
- Berdal, K.G. and A. Holst Jensen. 2001. Roundup Ready® soybean event-specific real-time quantitative PCR assay and estimation of the practical detection and quantification limits in GMO analyses. *Eur Food Res Technol*, **213**: 432-438.
- Berger, S., T. Mitchell Olds and H.U. Stotz. 2002. Local and differential control of vegetative storage protein expression in response to herbivore damage in *Arabidopsis thaliana*. *Physiol Plant*, **114**: 85-91.
- Bettini, P., S. Michelotti, D. Bindi, R. Giannini, M. Capuana and M. Buiatti. 2003. Pleiotropic effect of the insertion of the *Agrobacterium rhizogenes rolD* gene in tomato (*Lycopersicon esculentum* Mill.). *Theor Appl Genet*, **107**: 831-836.
- Bezier, A., B. Lambert and F. Baillieul. 2002. Study of defense-related gene expression in grapevine leaves and berries infected with *Botrytis cinerea*. *Eur J Plant Pathology*, **108**: 111-120.
- Blanco Portales, R., N. Medina Escobar, J.A. Lopez Raez, J.A. Gonzalez Reyes, J.M. Villalba, E. Moyano, J.L. Caballero and J. Muñoz Blanco. 2002. Cloning, expression and immunolocalization pattern of a cinnamyl alcohol dehydrogenase gene from strawberry (*Fragaria x ananassa* cv. Chandler). *J Exp Bot*, **53**: 1723-1734.
- Böhm, J., A. Hahn, R. Schubert, G. Bahnweg, N. Adler, J. Nechwatal, R. Oehlmann and W. Osswald. 1999. Real-time quantitative PCR: DNA determination in isolated spores of the mycorrhizal fungus *Glomus mosae* and monitoring of *Phytophthora infestans* and *Phytophthora citricola* in their respective host plants. *J Phytopathol*, **147**: 409-416.
- Bonaventure, G., J.J. Salas, M.R. Pollard and J.B. Ohlrogge. 2003. Disruption of the *FATB* gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell*, **15**: 1020-1033.
- Bovy, A., R. de Vos, M. Kemper, E. Schijlen, M.A. Pertejo, S. Muir, G. Collins, S. Robinson, M. Verhoeven, S. Hughes, C. Santos Buelga and A. van Tunen. 2002. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes *LC* and *Cl*. *Plant Cell*, **14**: 2509-2526.
- Brodmann, P.D., E.C. Ilg, H. Berthoud and A. Herrmann. 2002. Real-time quantitative polymerase chain reaction methods for four genetically modified maize varieties and maize DNA content in food. *J Aoac Int*, **85**: 646-653.
- Brown, R.L., K. Kazan, K.C. McGrath, D.J. Maclean and J.M. Manners. 2003. A role for the GCC-box in jasmonate-mediated activation of the *PDF1.2* gene of *Arabidopsis*. *Plant Physiol*, **132**: 1020-1032.
- Busch, M., A. Seuter and R. Hain. 2002. Functional analysis of the early steps of carotenoid biosynthesis in tobacco. *Plant Physiol*, **128**: 439-453.

- Bustin, S.A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol*, **25**: 169-193.
- Bustin, S.A. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol*, **29**: 23-39.
- Chae, H.S., F. Faure and J.J. Kieber. 2003. The *eto1*, *eto2*, and *eto3* mutations and cytokinin treatment increase ethylene biosynthesis in *Arabidopsis* by increasing the stability of ACS protein. *Plant Cell*, **15**: 545-559.
- Chang, H.S., M.L. Jones, G.M. Banowitz and D.G. Clark. 2003. Overproduction of cytokinins in petunia flowers transformed with P-SAG12-IPT delays corolla senescence and decreases sensitivity to ethylene. *Plant Physiol*, **132**: 2174-2183.
- Charrier, B., A. Champion, Y. Henry and M. Kreis. 2002. Expression profiling of the whole *Arabidopsis* Shaggy-like kinase multigene family by real-time reverse transcriptase-polymerase chain reaction. *Plant Physiol*, **130**: 577-590.
- Chen, I.P., U. Haehnel, L. Altschmied, I. Schubert and H. Puchta. 2003. The transcriptional response of *Arabidopsis* to genotoxic stress - a high-density colony array study (HDCA). *Plant J*, **35**: 771-786.
- Chien, A., D.B. Edgar and J.M. Trela. 1976. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J Bacteriol*, **127**: 1550-1557.
- Choi, D.W., E.M. Rodriguez and T.J. Close. 2002. Barley *Cbf3* gene identification, expression pattern, and map location. *Plant Physiol*, **129**: 1781-1787.
- Collakova, E. and D. DellaPenna. 2003. The role of homogentisate phytyltransferase and other tocopherol pathway enzymes in the regulation of tocopherol synthesis during abiotic stress. *Plant Physiol*, **133**: 930-940.
- Considine, M.J., D.O. Daley and J. Whelan. 2001. The expression of alternative oxidase and uncoupling protein during fruit ripening in mango. *Plant Physiol*, **126**: 1619-1629.
- Cullen, D.W., A.K. Lees, I.K. Toth and J.M. Duncan. 2001. Conventional PCR and real-time quantitative PCR detection of *Helminthosporium solani* in soil and on potato tubers. *Eur J Plant Pathology*, **107**: 387-398.
- Cullen, D.W., A.K. Lees, I.K. Toth and J.M. Duncan. 2002. Detection of *Colletotrichum coccodes* from soil and potato tubers by conventional and quantitative real-time PCR. *Plant Pathol*, **51**: 281-292.
- Dambrauskas, G., S.J. Aves, J.A. Bryant, D. Francis and H.J. Rogers. 2003. Genes encoding two essential DNA replication activation proteins, *Cdc6* and *Mem3*, exhibit very different patterns of expression in the tobacco BY-2 cell cycle. *J Exp Bot*, **54**: 699-706.
- Delanoy, M., M. Salmon, J. Kummert, E. Frison and P. Lepoivre. 2003. Development of real-time PCR for the rapid detection of episomal Banana streak virus (BSV). *Plant Dis*, **87**: 33-38.
- Dornelas, M.C., P. Wittich, I. von Recklinghausen, A. van Lammeren and M. Kreis. 1999. Characterization of three novel members of the *Arabidopsis* SHAGGY-related protein kinase (*ASK*) multigene family. *Plant Mol Biol*, **39**: 137-147.
- Ekman, D.R., W.W. Lorenz, A.E. Przybyla, N.L. Wolfe and J.F. Dean. 2003. SAGE analysis of transcriptome responses in *Arabidopsis* roots exposed to 2,4,6-trinitrotoluene. *Plant Physiol*, **133**: 1397-1406.
- Eun, A.J.C., M.L. Seoh and S.M. Wong. 2000. Simultaneous quantitation of two orchid viruses by the TaqMan® real-time RT-PCR. *J Virol Meth*, **87**: 151-160.
- Faccioli, P., M.S. Lagonigro, L. De Cecco, A.M. Stanca, R. Alberici and V. Terzi. 2002. Analysis of differential expression of barley ESTs during cold acclimatization using microarray technology. *Plant Biol*, **4**: 630-639.
- Ferre, F. 1992. Quantitative or semi-quantitative PCR: reality versus myth. *PCR Methods Appl*, **2**: 1-9.
- Filion, M., M. St Arnaud and S.H. Jabaji Hare. 2003a. Direct quantification of fungal DNA from soil substrate using real-time PCR. *J Microbiol Meth*, **53**: 67-76.
- Filion, M., M. St Arnaud and S.H. Jabaji Hare. 2003b. Quantification of *Fusarium solani* f. sp *phaseoli* in mycorrhizal bean plants and surrounding mycorrhizosphere soil using real-time polymerase chain reaction and direct isolations on selective media. *Phytopathology*, **93**: 229-235.
- Foucher, F., J. Morin, J. Courtiade, S. Cadioux, N. Ellis, M.J. Banfield and C. Rameau. 2003. DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. *Plant Cell*, **15**: 2742-2754.
- Fraaije, B.A., J.A. Butters, J.M. Coelho, D.R. Jones and D.W. Hollomon. 2002. Following the dynamics of strobilurin resistance in *Blumeria graminis* f.sp *tritici* using quantitative allele-specific real-time PCR measurements with the fluorescent dye SYBR Green I. *Plant Pathol*, **51**: 45-54.
- Fraaije, B.A., D.J. Lovell, J.M. Coelho, S. Baldwin and D.W. Hollomon. 2001. PCR-based assays to assess wheat varietal resistance to blotch (*Septoria tritici* and *Stagonospora nodorum*) and rust (*Puccinia striiformis* and *Puccinia recondita*) diseases. *Eur J Plant Pathology*, **107**: 905-917.
- Frederick, R.D., C.L. Snyder, G.L. Peterson and M.R. Bonde. 2002. Polymerase chain reaction assays for the detection and discrimination of the soybean rust pathogens *Phakopsora pachyrhizi* and *P-meibomiaae*. *Phytopathology*, **92**: 217-227.

- Freeman, W.M., S.J. Walker and K.E. Vrana. 1999. Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, **26**: 112-122, 124-115.
- German, M.A., M. Kandel Kfir, D. Swarzburg, T. Matsevitz and D. Granot. 2003. A rapid method for the analysis of zygosity in transgenic plants. *Plant Sci*, **164**: 183-187.
- Gilliland, G., S. Perrin, K. Blanchard and H.F. Bunn. 1990. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc Natl Acad Sci U S A*, **87**: 2725-2729.
- Goda, H., Y. Shimada, T. Asami, S. Fujioka and S. Yoshida. 2002. Microarray analysis of brassinosteroid-regulated genes in *Arabidopsis*. *Plant Physiol*, **130**: 1319-1334.
- Goncalves, M.C., M.M. Klerks, M. Verbeek, J. Vega and J. van den Heuvel. 2002. The use of molecular beacons combined with NASBA for the sensitive detection of Sugarcane yellow leaf virus. *Eur J Plant Pathology*, **108**: 401-407.
- Goto, D.B. and S. Naito. 2002. *AtMRD1* and *AtMRU1*, two novel genes with altered mRNA levels in the methionine over-accumulating mto1-1 mutant of *Arabidopsis thaliana*. *Plant Cell Physiol*, **43**: 923-931.
- Guescini, M., R. Pierleoni, F. Palma, S. Zeppa, L. Vallorani, L. Potenza, C. Sacconi, G. Giomaro and V. Stocchi. 2003. Characterization of the *Tuber borchii* nitrate reductase gene and its role in ectomycorrhizae. *Mol Genet Genomics*, **269**: 807-816.
- Hahnen, S., T. Joeris, F. Kreuzaler and C. Peterhansel. 2003. Quantification of photosynthetic gene expression in maize C-3 and C-4 tissues by real-time PCR. *Photosynth Res*, **75**: 183-192.
- Halterman, D.A., F.S. Wei and R.P. Wise. 2003. Powdery mildew-induced Mla mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiol*, **131**: 558-567.
- Hammond, J.P., M.J. Bennett, H.C. Bowen, M.R. Broadley, D.C. Eastwood, S.T. May, C. Rahn, R. Swarup, K.E. Woolaway and P.J. White. 2003. Changes in gene expression in *Arabidopsis* shoots during phosphate starvation and the potential for developing smart plants. *Plant Physiol*, **132**: 578-596.
- Harari-Steinberg, O., I. Ohad and D.A. Chamovitz. 2001. Dissection of the light signal transduction pathways regulating the two early light-induced protein genes in *Arabidopsis*. *Plant Physiol*, **127**: 986-997.
- Heid, C.A., J. Stevens, K.J. Livak and P.M. Williams. 1996. Real time quantitative PCR. *Genome Res*, **6**: 986-994.
- Hernandez, M., A. Ferrando, T. Esteve, P. Puigdomenech, S. Prat and M. Pla. 2003a. Real-time and conventional polymerase chain reaction systems based on the metallo-carboxypeptidase inhibitor gene for specific detection and quantification of potato and tomato in processed food. *J Food Protect*, **66**: 1063-1070.
- Hernandez, M., M. Pla, T. Esteve, S. Prat, P. Puigdomenech and A. Ferrando. 2003b. A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard® based on the 3'-transgene integration sequence. *Transgenic Res*, **12**: 179-189.
- Hernandez, M., A. Rio, T. Esteve, S. Prat and M. Pla. 2001. A rapeseed-specific gene, Acetyl-CoA carboxylase, can be used as a reference for qualitative and real-time quantitative PCR detection of transgenes from mixed food samples. *J Agr Food Chem*, **49**: 3622-3627.
- Heuser, T. and W. Zimmer. 2002. Quantitative analysis of phytopathogenic Ascomycota on leaves of pedunculate oaks (*Quercus robur* L.) by real-time PCR. *FEMS Microbiol Lett*, **209**: 295-299.
- Higgins, J.A., M. Cooper, L. Schroeder Tucker, S. Black, D. Miller, J.S. Karns, E. Manthey, R. Breeze and M.L. Perdue. 2003. A field investigation of *Bacillus anthracis* contamination of US Department of Agriculture and other Washington, DC, buildings during the anthrax attack of October 2001. *Appl Environ Microbiol*, **69**: 593-599.
- Higuchi, R., C. Fockler, G. Dollinger and R. Watson. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y)*, **11**: 1026-1030.
- Hird, H., J. Powell, M.L. Johnson and S. Oehlschlager. 2003. Determination of percentage of RoundUp Ready® soya in soya flour using real-time polymerase chain reaction: Interlaboratory study. *J Aoac Int*, **86**: 66-71.
- Hohne, M., C.R. Santisi and R. Meyer. 2002. Real-time multiplex PCR: An accurate method for the detection and quantification of 35S-CaMV promoter in genetically modified maize-containing food. *Eur Food Res Technol*, **215**: 59-64.
- Holmberg, N., M. Harker, C.L. Gibbard, A.D. Wallace, J.C. Clayton, S. Rawlins, A. Hellyer and R. Safford. 2002. Sterol C-24 methyltransferase type 1 controls the flux of carbon into sterol biosynthesis in tobacco seed. *Plant Physiol*, **130**: 303-311.
- Ingham, D.J., S. Beer, S. Money and G. Hansen. 2001. Quantitative real-time PCR assay for determining transgene copy number in transformed plants. *Biotechniques*, **31**: 132,134,136-140.
- Ito, S., Y. Ushijima, T. Fujii, S. Tanaka, M. Kameya-Iwaki and e. al. 1998. Detection of viable cells of *Ralstonia solanacearum* in soil using semiselective medium and a PCR technique. *J Phytopathol*, **146**: 379-384.
- Itoh, R., M. Fujiwara, N. Nagata and S. Yoshida. 2001. A chloroplast protein homologous to the eubacterial topological specificity factor MinE plays a role in chloroplast division. *Plant Physiol*, **127**: 1644-1655.

- Jacobs, A.K., V. Lipka, R.A. Burton, R. Panstruga, N. Strizhov, P. Schulze-Lefert and G.B. Fincher. 2003. An *Arabidopsis* callose synthase, *GSL5*, is required for wound and papillary callose formation. *Plant Cell*, **15**: 2503-2513.
- Jakab, G., A. Manrique, L. Zimmerli, J.P. Metraux and B. Mauch Mani. 2003. Molecular characterization of a novel lipase-like pathogen-inducible gene family of *Arabidopsis*. *Plant Physiol*, **132**: 2230-2239.
- Jones, L., F. Ratcliff and D.C. Baulcombe. 2001. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr Biol*, **11**: 747-757.
- Jones, P.G., D. Allaway, D.M. Gilmour, C. Harris, D. Rankin, E.R. Retzel and C.A. Jones. 2002. Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. *Planta*, **216**: 255-264.
- Karsai, A., S. Muller, S. Platz and M.T. Hauser. 2002. Evaluation of a homemade SYBR® Green I reaction mixture for real-time PCR quantification of gene expression. *Biotechniques*, **32**: 790-792, 794-796.
- Kingsnorth, C.S., A.J. Kingsnorth, P.A. Lyons, D.M. Chwarszczynska and M.J.C. Asher. 2003. Real-time analysis of *Polymyxa betae* GST expression in infected sugar beet. *Mol Plant Pathol*, **4**: 171-176.
- Klok, E.J., I.W. Wilson, D. Wilson, S.C. Chapman, R.M. Ewing, S.C. Somerville, W.J. Peacock, R. Dolferus and E.S. Dennis. 2002. Expression profile analysis of the low-oxygen response in *Arabidopsis* root cultures. *Plant Cell*, **14**: 2481-2494.
- Kuribara, H., Y. Shindo, T. Matsuoka, K. Takubo, S. Futo, N. Aoki, T. Hirao, H. Akiyama, Y. Goda, M. Toyoda and A. Hino. 2002. Novel reference molecules for quantitation of genetically modified maize and soybean. *J Aoac Int*, **85**: 1077-1089.
- Kursteiner, O., I. Dupuis and C. Kuhlemeier. 2003. The pyruvate decarboxylase1 gene of *Arabidopsis* is required during anoxia but not other environmental stresses. *Plant Physiol*, **132**: 968-978.
- Lacomme, C., K. Hrubikova and I. Hein. 2003. Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. *Plant J*, **34**: 543-553.
- Lake, V. and R.D. Willows. 2003. Rapid extraction of RNA and analysis of transcript levels in *Chlamydomonas reinhardtii* using real-time RT-PCR: Magnesium chelatase *chlH*, *chlD* and *chlI* gene expression. *Photosynth Res*, **77**: 69-76.
- Lees, A.K., D.W. Cullen, L. Sullivan and M.J. Nicolson. 2002. Development of conventional and quantitative real-time PCR assays for the detection and identification of *Rhizoctonia solani* AG-3 in potato and soil. *Plant Pathol*, **51**: 293-302.
- Li, W.B., W.D. Pria, P.M. Lacava, X. Qin and J.S. Hartung. 2003. Presence of *Xylella fastidiosa* in sweet orange fruit and seeds and its transmission to seedlings. *Phytopathology*, **93**: 953-958.
- Liu, J.P., B. Cong and S.D.C.-R.-A.-B. Tanksley. 2003. Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the cloned quantitative trait locus fw2.2 controls fruit size. *Plant Physiol*, **132**: 292-299.
- Maguire, T.L., S. Grimmond, A. Forrest, I. Iturbe Ormaetxe, K. Meksem and P. Gresshoff. 2002. Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis. *J Plant Physiol*, **159**: 1361-1374.
- Maruyama-Nakashita, A., E. Inoue, A. Watanabe-Takahashi, T. Yarnaya and H. Takahashi. 2003. Transcriptome profiling of sulfur-responsive genes in *Arabidopsis* reveals global effects of sulfur nutrition on multiple metabolic pathways. *Plant Physiol*, **132**: 597-605.
- McCartney, H.A., S.J. Foster, B.A. Fraaije and E. Ward. 2003. Molecular diagnostics for fungal plant pathogens. *Pest Manag Sci*, **59**: 129-142.
- McGinnis, K.M., S.G. Thomas, J.D. Soule, L.C. Strader, J.M. Zale, T.P. Sun and C.M. Steber. 2003. The *Arabidopsis* *SLEEPY1* gene encodes a putative F-box subunit of an SCF E3 ubiquitin ligase. *Plant Cell*, **15**: 1120-1130.
- Mills, D. and B.W. Russell. 2003. Parameters for specific detection of *Clavibacter michiganensis* subsp *sepedonicus* in potato stems and tubers by multiplexed PCR-ELISA. *Am J Potato Res*, **80**: 223-233.
- Mladek, C., K. Guger and M.T. Hauser. 2003. Identification and characterization of the *ARIADNE* gene family in *Arabidopsis*. A group of putative E3 ligases. *Plant Physiol*, **131**: 27-40.
- Montrichard, F., M. Renard, F. Alkhalfioui, F.D. Duval and D. Macherel. 2003. Identification and differential expression of two thioredoxin h isoforms in germinating seeds from pea. *Plant Physiol*, **132**: 1707-1715.
- Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, **51**: 263-273.
- Mullis, K.B. and F.A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*, **155**: 335-350.
- Mumford, R.A., K. Walsh, I. Barker and N. Boonham. 2000. Detection of potato mop top virus and tobacco rattle virus using a multiplex real-time fluorescent reverse-transcription polymerase chain reaction assay. *Phytopathology*, **90**: 448-453.

- Murcha, M.W., R. Lister, A.Y.Y. Ho and J. Whelan. 2003. Identification, expression, and import of components 17 and 23 of the inner mitochondrial membrane translocase from *Arabidopsis*. *Plant Physiol*, **131**: 1737-1747.
- Nap, J.P., P.L.J. Metz, M. Escaler and A.J. Conner. 2003. The release of genetically modified crops into the environment - Part I. Overview of current status and regulations. *Plant J*, **33**: 1-18.
- Nicolaisen, M. 2003. Partial molecular characterization of Dahlia mosaic virus and its detection by PCR. *Plant Dis*, **87**: 945-948.
- Nolan, K.E., R.R. Irwanto and R.J. Rose. 2003. Auxin up-regulates MtSERK1 expression in both *Medicago truncatula* root-forming and embryogenic cultures. *Plant Physiol*, **133**: 218-230.
- Ogawa, M., A. Hanada, Y. Yamauchi, A. Kuwalhara, Y. Kamiya and S. Yamaguchi. 2003. Gibberellin biosynthesis and response during *Arabidopsis* seed germination. *Plant Cell*, **15**: 1591-1604.
- Oliveira, A.C., M.A. Vallim, C.P. Semighini, W.L. Araujo, G.H. Goldman and M.A. Machado. 2002. Quantification of *Xylella fastidiosa* from citrus trees by real-time polymerase chain reaction assay. *Phytopathology*, **92**: 1048-1054.
- Orsel, M., A. Krapp and F. Daniel Vedele. 2002. Analysis of the *NRT2* nitrate transporter family in *Arabidopsis*. Structure and gene expression. *Plant Physiol*, **129**: 886-896.
- Panchuk, II, R.A. Volkov and F. Schoffl. 2002. Heat stress- and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in *Arabidopsis*. *Plant Physiol*, **129**: 838-853.
- Peeters, N.M. and M.R. Hanson. 2002. Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA*, **8**: 497-511.
- Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*, **29**: e45.
- Philippar, K., K. Buchsenschutz, M. Abshagen, I. Fuchs, D. Geiger, B. Lacombe and R. Hedrich. 2003. The K<sup>+</sup> channel KZM1 mediates potassium uptake into the phloem and guard cells of the C-4 grass *Zea mays*. *J Biol Chem*, **278**: 16973-16981.
- Pietila, J., Q.S. He, J. Oksi and M.K. Viljanen. 2000. Rapid differentiation of *Borrelia garinii* from *Borrelia afzelii* and *Borrelia burgdorferi* sensu stricto by LightCycler fluorescence melting curve analysis of a PCR product of the *recA* gene. *Journal of Clinical Microbiology*, **38**: 2756-2759.
- Puthoff, D.P., D. Nettleton, S.R. Rodermeil and T.J. Baum. 2003. *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. *Plant J*, **33**: 911-921.
- Qi, M. and Y.N. Yang. 2002. Quantification of *Magnaporthe grisea* during infection of rice plants using real-time polymerase chain reaction and northern blot/phosphoimaging analyses. *Phytopathology*, **92**: 870-876.
- Raeymaekers, L. 1993. Quantitative PCR: theoretical considerations with practical implications. *Anal Biochem*, **214**: 582-585.
- Rappolee, D.A., D. Mark, M.J. Banda and Z. Werb. 1988. Wound macrophages express TGF- $\alpha$  and other growth factors in vivo: analysis by mRNA phenotyping. *Science*, **241**: 708-712.
- Reintanz, B., A. Szyroki, N. Ivashikina, P. Ache, M. Godde, D. Becker, K. Palme and R. Hedrich. 2002. AtKC1, a silent *Arabidopsis* potassium channel alpha-subunit modulates root hair K<sup>+</sup> influx. *Proc Nat Acad Sci USA*, **99**: 4079-4084.
- Rider, S.D., J.T. Henderson, R.E. Jerome, H.J. Edenberg, J. Romero Severson and J. Ogas. 2003. Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis*. *Plant J*, **35**: 33-43.
- Roberts, C.A., R.G. Dietzgen, L.A. Heelan and D.J. Maclean. 2000. Real-time RT-PCR fluorescent detection of tomato spotted wilt virus. *J Virol Meth*, **88**: 1-8.
- Ronning, S.B., M. Vaitilingom, K.G. Berdal and A. Holst Jensen. 2003. Event specific real-time quantitative PCR for genetically modified Bt11 maize (*Zea mays*). *Eur Food Res Technol*, **216**: 347-354.
- de Ruijter, N.C.A., J. Verhees, W. van Leeuwen and A.R. van der Krol. 2003. Evaluation and comparison of the GUS, LUC and GFP reporter system for gene expression studies in plants. *Plant Biol*, **5**: 103-115.
- Saiki, R.K., S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich and N. Arnheim. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**: 1350-1354.
- Salmon, M.A., M. Vendrame, J. Kummert and P. Lepoivre. 2002. Detection of apple chlorotic leaf spot virus using a 5' nuclease assay with a fluorescent 3' minor groove binder-DNA probe. *J Virol Meth*, **104**: 99-106.
- Sandberg, M., L. Lundberg, M. Ferm and I.M. Yman. 2003. Real time PCR for the detection and discrimination of cereal contamination in gluten free foods. *Eur Food Res Technol*, **217**: 344-349.
- Schaad, N.W. and R.D. Frederick. 2002. Real-time PCR and its application for rapid plant disease diagnostics. *Can J Plant Pathol*, **24**: 250-258.

- Schaad, N.W., R.D. Frederick, A. Shalev, J. Shaw, W.L. Schneider, R. Hickson, M.D. Petrillo and D.G. Luster. 2003. Advances in molecular-based diagnostics in meting crop biosecurity and phytosanitary issues. *Ann Rev Phytopathol*, **41**: 305-324.
- Schaad, N.W., D. Opgenorth and P. Gauth. 2002. Real-time polymerase chain reaction for one-hour on-site diagnosis of Pierce's disease of grape in early season asymptomatic vines. *Phytopathology*, **92**: 721-728.
- Schena, L., F. Nigro and A. Ippolito. 2002. Identification and detection of *Rosellinia necatrix* by conventional and real-time Scorpion-PCR. *Eur J Plant Pathology*, **108**: 355-366.
- Schenk, P.M., K. Kazan, J.M. Manners, J.P. Anderson, R.S. Simpson, I.W. Wilson, S.C. Somerville and D.J. Maclean. 2003. Systemic gene expression in *Arabidopsis* during an incompatible interaction with *Alternaria brassicicola*. *Plant Physiol*, **132**: 999-1010.
- Schmidt, M.A. and W.A. Parrott. 2001. Quantitative detection of transgenes in soybean [*Glycine max* (L.) Merrill] and peanut (*Arachis hypogaea* L.) by real-time polymerase chain reaction. *Plant Cell Rep*, **20**: 422-428.
- Schnerr, H., L. Niessen and R.F. Vogel. 2000. Real time detection of the *tri5* gene in *Fusarium* species by LightCycler<sup>TM</sup>-PCR using SYBR<sup>®</sup>-Green I for continuous fluorescence monitoring. *Int J Food Microbiol*, **71**: 53-61.
- Schoen, C.D., D. Knorr and G. Leone. 1996. Detection of potato leafroll virus in dormant potato tubers by immunocapture and a fluorogenic 5' nuclease RT-PCR assay. *Phytopathology*, **86**: 993-999.
- Schubert, R., S. Raidl, R. Funk, G. Bahnweg, G. Muller Starck and R. Agerer. 2003. Quantitative detection of agar-cultivated and rhizotron-grown *Piloderma croceum* Erikss. and Hjortst. by ITS1-based fluorescent PCR. *Mycorrhiza*, **13**: 159-165.
- Shen, G.A., Y.Z. Pang, C.F. Lin, C. Wei, X.Y. Qian, L.Z. Jiang, X.L. Du, K.G. Li, K. Attia and J.S. Yang. 2003. Cloning and characterization of a novel *Hsp100/Clp* gene (*osClpD*) from *Oryza sativa*. *DNA Sequence*, **14**: 285-293.
- Shimada, Y., H. Goda, A. Nakamura, S. Takatsuto, S. Fujioka and S. Yoshida. 2003. Organ-specific expression of brassinosteroid-biosynthetic genes and distribution of endogenous brassinosteroids in *Arabidopsis*. *Plant Physiol*, **131**: 287-297.
- Skaf, J.S., M.H. Schultz, H. Hirata and G.A. de Zoeten. 2000. Mutational evidence that the VPg is involved in the replication and not the movement of Pea enation mosaic virus-1. *J Gen Virol*, **81**: 1103-1109.
- Smith, R.D., B. Brown, P. Ikononi and A.N. Schechter. 2003. Exogenous reference RNA for normalization of real-time quantitative PCR. *Biotechniques*, **34**: 88-91.
- Song, P., C.Q. Cai, M. Skokut, B.D. Kosegi and J.F. Petolino. 2002. Quantitative real-time PCR as a screening tool for estimating transgene copy number in WHISKERS<sup>TM</sup>-derived transgenic maize. *Plant Cell Rep*, **20**: 948-954.
- Spackman, E., D.A. Senne, L.L. Bulaga, S. Trock and D.L. Suarez. 2003. Development of multiplex real-time RT-PCR as a diagnostic tool for avian influenza. *Avian Dis*, **47 Suppl. S**: 1087-1090.
- Stasolla, C., L. van Zyl, U. Egertsdotter, D. Craig, W.B. Liu and R.R. Sederoff. 2003. The effects of polyethylene glycol on gene expression of developing white spruce somatic embryos. *Plant Physiol*, **131**: 49-60.
- Suzuki, H., T. Nakayama, K. Yonekura Sakakibara, Y. Fukui, N. Nakamura, M. Yamaguchi, Y. Tanaka, T. Kusumi and T. Nishino. 2002. cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A : anthocyanidin 3-O-glucoside-6"-O-malonyltransferase from dahlia flowers. *Plant Physiol*, **130**: 2142-2151.
- Svensson, A.S., F.I. Johnsson, I.M. Moller and A.G. Rasmusson. 2002. Cold stress decreases the capacity for respiratory NADH oxidation in potato leaves. *Febs Lett*, **517**: 79-82.
- Tan, B.C., L.M. Joseph, W.T. Deng, L.J. Liu, Q.B. Li, K. Cline and D.R. McCarty. 2003. Molecular characterization of the *Arabidopsis* 9-cis epoxy-carotenoid dioxygenase gene family. *Plant J*, **35**: 44-56.
- Taverniers, I., P. Windels, E. Van Bockstaele and M. De Loose. 2001. Use of cloned DNA fragments for event-specific quantification of genetically modified organisms in pure and mixed food products. *Eur Food Res Technol*, **213**: 417-424.
- Taylor, G.R. and P. Robinson. 1998. The polymerase chain reaction: from functional genomics to high-school practical classes. *Curr Opin Biotechnol*, **9**: 35-42.
- Teramoto, H., A. Nakamori, J. Minagawa and T. Ono. 2002. Light-intensity-dependent expression of *Lhc* gene family encoding light-harvesting chlorophyll-a/b proteins of photosystem II in *Chlamydomonas reinhardtii*. *Plant Physiol*, **130**: 325-333.
- Terzi, V., M. Malnati, M. Barbanera, A.M. Stanca and P. Faccioli. 2003. Development of analytical systems based on real-time PCR for *Triticum* species-specific detection and quantitation of bread wheat contamination in semolina and pasta. *J Cereal Sci*, **38**: 87-94.
- Thomas, C., D. Meyer, M. Wolff, C. Hember, M. Alioua and A. Steinmetz. 2003. Molecular characterization and spatial expression of the sunflower *ABPI* gene. *Plant Mol Biol*, **52**: 1025-1036.

- Tierens, K., B.P.H. Thomma, M. Brouwer, J. Schmidt, K. Kistner, A. Porzel, B. Mauch Mani, B.P.A. Cammue and W.F. Broekaert. 2001. Study of the role of antimicrobial glucosinolate-derived isothiocyanates in resistance of arabidopsis to microbial pathogens. *Plant Physiol*, **125**: 1688-1699.
- Ullanat, R. and C. Jayabaskaran. 2002. Distinct light-, cytokinin- and tissue-specific regulation of calcium dependent protein kinase gene expression in cucumber (*Cucumis sativus*). *Plant Sci*, **162**: 153-163.
- Vaitilingom, M., H. Pijnenburg, F. Gendre and P. Brignon. 1999. Real-time quantitative PCR detection of genetically modified maximizer maize and roundup ready soybean in some representative foods. *J Agr Food Chem*, **47**: 5261-5266.
- van Wees, S.C.M., H.S. Chang, T. Zhu and J. Glazebrook. 2003. Characterization of the early response of *Arabidopsis* to *Alternaria brassicicola* infection using expression profiling. *Plant Physiol*, **132**: 606-617.
- Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*, **3**: research0034.0031-research0034.0011.
- Volkov, R.A., Panchuk, II and F. Schoffl. 2003. Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR. *J Exp Bot*, **54**: 2343-2349.
- Wang, R.C., M. Okamoto, X.J. Xing and N.M. Crawford. 2003. Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiol*, **132**: 556-567.
- Weller, S.A., J.G. Elphinstone, N.C. Smith, N. Boonham and D.E. Stead. 2000. Detection of *Ralstonia solanacearum* strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan) assay. *Appl Environ Microbiol*, **66**: 2853-2858.
- Weller, S.A., S.A. Simpkins, D.E. Stead, A. Kurdziel, H. Hird and R.J. Weekes. 2002. Identification of *Agrobacterium spp.* present within *Brassica napus* seed by TaqMan PCR - implications for GM screening procedures. *Arch Microbiol*, **178**: 338-343.
- Weller, S.A. and D.E. Stead. 2002. Detection of root mat associated *Agrobacterium* strains from plant material and other sample types by post-enrichment TaqMan PCR. *J Appl Microbiol*, **92**: 118-126.
- Wilhelm, J. and A. Pingoud. 2003. Real-time polymerase chain reaction. *ChemBiochem*, **4**: 1120-1128.
- Windels, P., S. Bertrand, A. Depicker, W. Moens, E. Bockstaele and M. Loose. 2003. Qualitative and event-specific PCR real-time detection methods for StarLink maize. *Eur Food Res Technol*, **216**: 259-263.
- Winton, L.M., D.K. Manter, J.K. Stone and E.A. Hansen. 2003. Comparison of biochemical, molecular, and visual methods to quantify *Phaeocryptopus gaeumannii* in Douglas-Fir foliage. *Phytopathology*, **93**: 121-126.
- Wittwer, C.T., M.G. Herrmann, A.A. Moss and R.P. Rasmussen. 1997a. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, **22**: 130-131, 134-138.
- Wittwer, C.T., K.M. Ririe, R.V. Andrew, D.A. David, R.A. Gundry and U.J. Balis. 1997b. The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques*, **22**: 176-181.
- Wurmbach, E., T. Yuen and S. Sealfon. 2003. Focused microarray analysis. *Methods*, **31**: 306-316.
- Yokoyama, R. and K. Nishitani. 2001. A comprehensive expression analysis of all members of a gene family encoding cell-wall enzymes allowed us to predict cis-regulatory regions involved in cell-wall construction in specific organs of *Arabidopsis*. *Plant Cell Physiol*, **42**: 1025-1033.
- Zeitler, R., M. Pietsch and H.U. Waiblinger. 2002. Validation of real-time PCR methods for the quantification of transgenic contaminations in rape seed. *Eur Food Res Technol*, **214**: 346-351.
- Zheng, Z.L., M. Nafisi, A. Tam, H. Li, D.N. Crowell, S.N. Chary, J.I. Schroeder, J.J. Shen and Z.B. Yang. 2002. Plasma membrane-associated ROP10 small GTPase is a specific negative regulator of abscisic acid responses in *Arabidopsis*. *Plant Cell*, **14**: 2787-2797.

## **Real-time PCR: what relevance to plant studies?**

*Journal of Experimental Botany*. 2004. Sous presse

Claire Gachon, Annaïck Mingam, Bénédicte Charrier

Institut de Biotechnologie des Plantes UMR CNRS 8618  
Université Paris-Sud  
91405 Orsay cedex. France.





## REVIEW ARTICLE

# Real-time PCR: what relevance to plant studies?

Claire Gachon, Annaïck Mingam and Bénédicte Charrier\*

*Institut de Biotechnologie des Plantes, UMR CNRS 8618, Université Paris-Sud, F-91405 Orsay cedex, France*

Received 19 January 2004; Accepted 22 April 2004

## Abstract

The appearance of genetically modified organisms on the food market a few years ago, and the demand for more precise and reliable techniques to detect foreign (transgenic or pathogenic) DNA in edible plants, have been the driving force for the introduction of real-time PCR techniques in plant research. This was followed by numerous fundamental research applications aiming to study the expression profiles of endogenous genes and multigene families. Since then, the interest in this technique in the plant scientist community has increased exponentially. This review describes the technical features of quantitative real-time PCR that are especially relevant to plant research, and summarizes its present and future applications.

Key words: Expression, fluorochrome, genetically modified organism, Molecular Beacon, pathogen, Scorpion<sup>TM</sup>, SYBR-green<sup>®</sup>, TaqMan<sup>®</sup>, technique, transgene.

## Introduction

Real-time PCR differs from classical PCR by the measurement of the amplified PCR product at each cycle throughout the PCR reaction. In practice, a video camera records the light emitted by a fluorochrome incorporated into the newly synthesized PCR product. Thus, real-time PCR allows the amplification to be followed in real-time during the exponential phase of the run, and thus allows the amount of starting material to be determined precisely. Contrary to end-point PCR techniques, the result is independent from the plateau corresponding to the saturation of the reaction, the latter leading to inaccurate quantification. Figure 1 shows the main principles of the real-time PCR process, while the details and requirements necessary to obtain

reliable data have been reviewed by Freeman *et al.* (1999) and Bustin (2000, 2002). Besides being an alternative to some well-established laboratory techniques, real-time PCR has a number of features which makes it the choice for several types of study. Compared with the other techniques presently available, it allows the detection of a given nucleic acid target in a rapid, specific and very sensitive way. In addition, it affords the absolute quantification of the initial target. To date, the reliability of real-time PCR has never been questioned.

## Relevant features of real-time PCR

### Rapidity

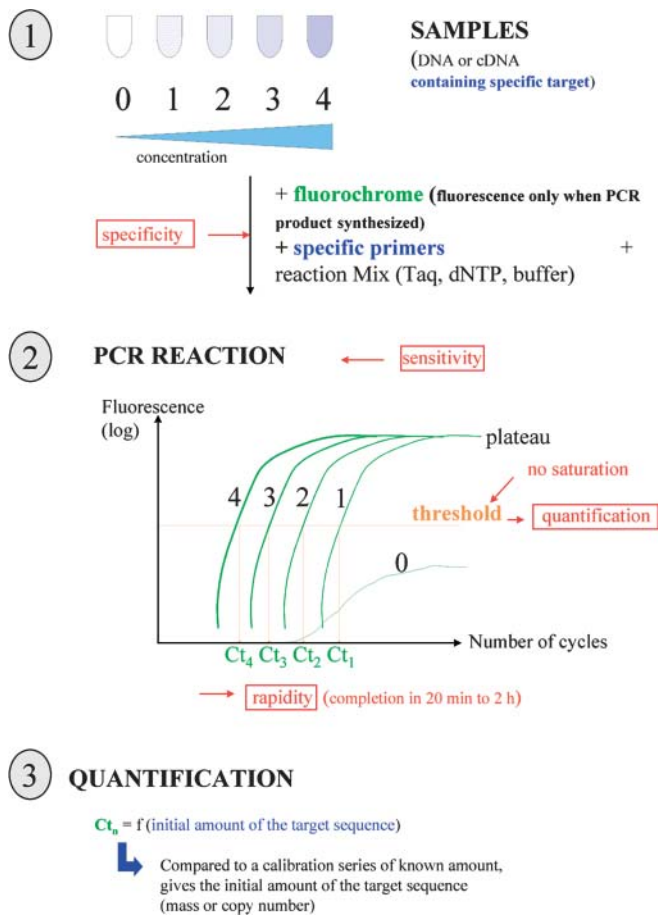
Compared with classical PCR, one of the main advantages of real-time PCR is its rapidity to provide reliable data. Typically, the time of a whole real-time PCR run ranges from 20 min to 2 h. Indeed, the time needed to shift temperature is a major limiting factor responsible for the duration of a classical PCR experiment. The LightCycler<sup>TM</sup> PCR machine (Roche) uses capillaries instead of tubes, which are heated by light instead of a heating block. As a result, the time necessary to heat the PCR mixture is considerably reduced (from 15 s to 1–2 s). In addition, recording the amplification in real-time avoids collecting samples at different steps of the PCR experiment, making the process less tedious and time-consuming. Moreover, some machines accommodate 384 well plates and can process queuing plates over 24 h non-stop (Wurmbach *et al.*, 2003), which might be a determining advantage for high throughput studies, or if rapid sample processing is required (Schnerr *et al.*, 2001).

### Sensitivity

Real-time PCR provides a high sensitivity for the detection of DNA or RNA due to a combination of the amplification

\* To whom correspondence should be addressed. Fax: +33 1 6915 3425. E-mail: charrier@ibp.u-psud.fr

Abbreviations: FRET, fluorescence resonance energy transfer; GMO, genetically modified organism; GMP, genetically modified plant; PCR, polymerase chain reaction.



**Fig. 1.** Summary of the principles of real-time quantitative PCR. (1) Real-time PCR is used to measure accurately the different amounts of a target gene product present in independent samples. In the case of real-time RT-PCR, the samples are cDNA previously reverse-transcribed from RNA preparations. (2) Each sample undergoes a PCR amplification together with a fluorochrome present in the reaction mixture, which fluoresces only when the specific product is recognized or synthesized (Fig. 2). A fluorescence threshold is determined in the linear part of the curve, corresponding to the phase with the best efficiency of amplification. A 'Ct' is defined as the number of cycles necessary to reach this threshold of fluorescence.  $C_{t_i} = f(\log_{10}$  initial concentration of the specific target). (3) A calibration curve is drawn from parallel runs using known initial amounts of the specific target. This allows the calculation of the amount of the target product present in a given sample of interest.

performed by the PCR step and the system of detection (Bustin, 2000). It is therefore a very convenient technique for studies with a limited amount of starting material (Bago *et al.*, 2002), or for assessing the expression of a high number of genes from minute quantities of RNA. The detection is based on the measurement of the fluorescence emitted by probes incorporated into the newly formed PCR product, or alternatively released into the buffer during the amplification of the PCR product. Intercalating agents and fluorogenic probes are the two main types of molecules currently used to detect PCR amplification in real-time. The first intercalating agent used was ethidium bromide, but in 1997, Wittwer *et al.* (1997a, b) proposed replacing it with the SYBRgreen<sup>®</sup> molecule, because of its higher affinity

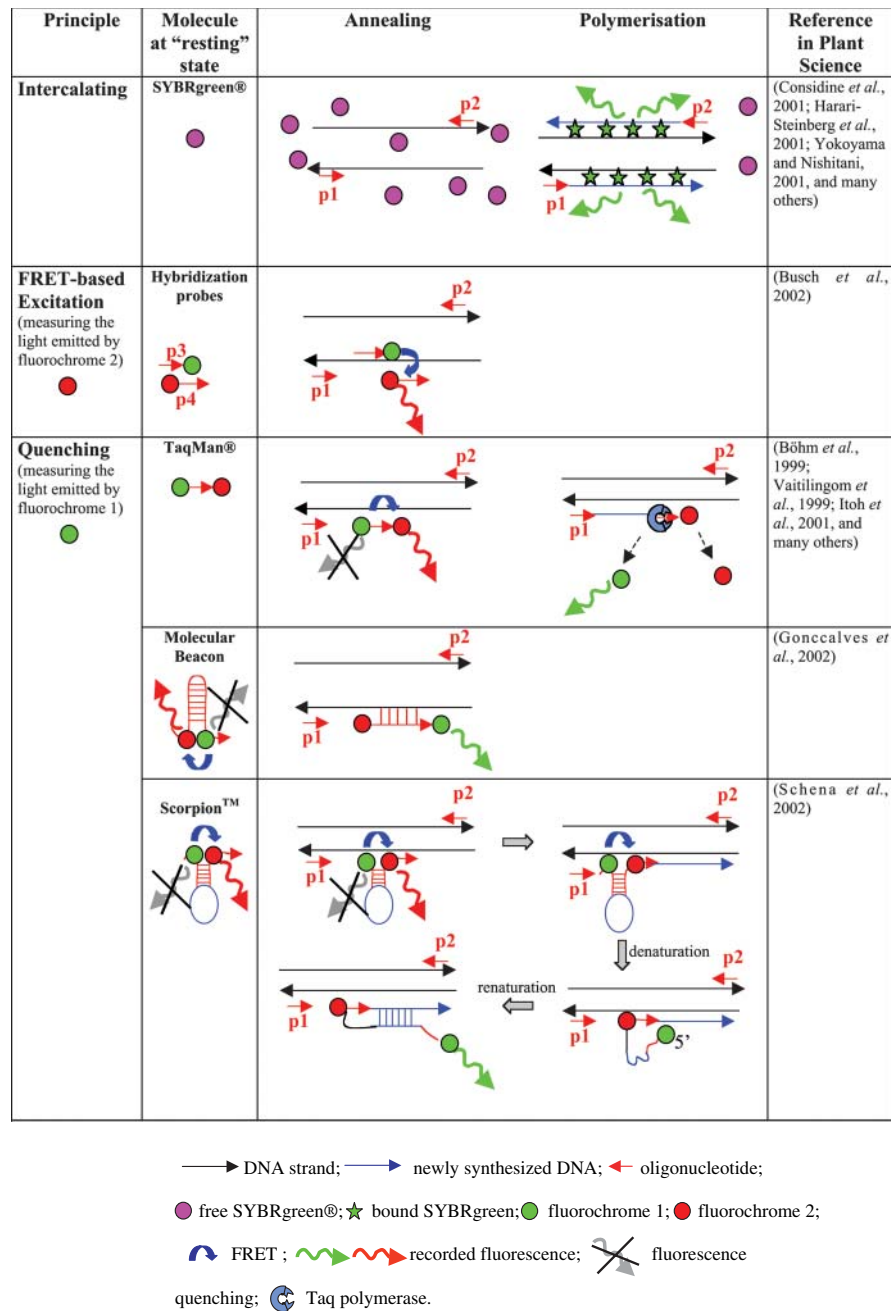
for double-stranded DNA. As intercalating agents bind regardless of the nucleotide nature, they can be used for any type of sequence. This is an economical advantage for a laboratory testing a large number of genes. However, a disadvantage of SYBRgreen<sup>®</sup> is that it is equally incorporated into every amplicon, and should unspecific sequences be amplified, the signal measured would correspond to both non-specific and specific products, thereby compromising the accurate quantification of the latter (see section 'Specificity' for further developments on this issue).

In order to bypass this potential problem, intercalating agents can be replaced by labelled oligonucleotides or probes, which specifically bind to the target sequence. This technology relies on the use of probes labelled with two different fluorochromes, one of which, when excited, is able to transfer its energy to the other via Fluorescence Resonance Energy Transfer (FRET). This non-radiative energy transfer only occurs if the two molecules are in close proximity to each other (a few nanometres). Depending on the proximity of the second fluorochrome, the first one may either emit light or transfer its energy to the second, which in turn fluoresces. Thus, bringing the two fluorochromes in close proximity to each other results in the fluorescence quenching of the first one, and fluorescence emission of the second one.

As the fluorescence from the emitting fluorochrome increases proportionally with the amount of newly synthesized DNA, both effects can be recorded to follow the amplification of the target DNA. Hence, several strategies have been developed, all of which rely on placing them in the vicinity of each other (excitation) or conversely ensuring their separation (quenching) during the amplification phase. So far, FRET-mediated excitation has rarely been used in plant studies (Busch *et al.*, 2002), and its use, which requires four oligonucleotides, should be limited to studies requiring a very high level of specificity. The application of quenching systems has been more common in plant studies, being initiated by the use of the TaqMan<sup>®</sup> probes, followed by the Molecular Beacons and Scorpion<sup>™</sup> probes. As they require the design of a labelled oligonucleotide specific for each sequence, they are economically relevant only if many experiments are to be performed on the same target. Figure 2 illustrates the main strategies currently developed and used in plant studies.

### Specificity

Surprisingly, in a study carried out on four pea thioredoxin h (TRXh) encoding genes, Montrichard *et al.* (2003) noticed that real-time PCR yielded weaker signals than expected from northern blot analyses. This observation was explained by a cross-hybridization of the probe to the RNA encoding another isoform during the northern blot procedure. Indeed, in contrast to techniques requiring the hybridization of nucleic acids several hundreds base pairs long, such as cDNA-based microarray and northern blotting, short oligonucleotide-mediated real-time PCR guarantees a high



**Fig. 2.** Main strategies for real-time specific detection of amplified PCR products. Note that increasing the number of oligonucleotides results in higher specificity.

specificity in the detection of the target sequence. In fact, specificity is achieved by the use of two target sequence-specific oligonucleotides, and this can be enhanced by increasing the number of oligonucleotides nested within the initial amplification product. In this respect, FRET-mediated probes seem to ensure a higher specificity than SYBRgreen® (Shimada *et al.*, 2003). In any case, specificity of the process can be checked after completion of the PCR run, by testing the nature of the amplified product with gel electrophoresis, melting curves, and sequencing data.

### Quantification

Most importantly, the quantification range of real-time PCR is up to seven orders of magnitude as originally illustrated by Higuchi *et al.* (1993), Heid *et al.* (1996), and more recently in plants by Charrier *et al.* (2002), Filion *et al.* (2003), and Hernández *et al.* (2003a). This results from the capacity of this technique to calculate, for every sample within an extremely low to high concentrations range, the number of cycles necessary to reach the Ct (see Fig. 1 for definition). The absolute amount of the target is calculated

from a calibration curve. Alternatively, a relative quantification can be deduced considering Ct differences between samples and standards as nicely illustrated by Bovy *et al.* (2002), and improved by Pfaffl (2001).

Basically, real-time quantitative PCR may be used for quantifying DNA or RNA abundance, leading to two major types of applications: foreign DNA (e.g. transgenes or contaminating micro-organisms) detection and quantification, and gene expression studies.

## Detection and quantification of foreign DNA

### *Quantification of pathogenic or symbiotic micro-organisms associated with plants*

Real-time PCR assays aiming at quantifying the level of plant infection by a pathogen have been increasing for the last few years, since the first report by Böhm *et al.* (1999). Most of them rely on the relative quantification of two specific plant and pathogen DNA sequences. They are faster, more specific and more sensitive when compared with traditional protocols based on symptom recording or on conidiophore or colony counting (Winton *et al.*, 2003), and, most importantly, may be transposed to virtually every pathosystem. For those reasons, they are being widely used for the diagnosis of diseases in the field and for applied purposes. For instance, seed potatoes cannot be sold in the EU unless they are devoid of the potato brown rot agent *Ralstonia solanacearum* (Council directive 2002/56/EC). Classical detection methods require a labour-intensive culture and pathogenicity test on tomato seedlings. However, real-time PCR has been shown to enable the quantitative detection of *R. solanacearum* in a rapid and reliable manner, thus providing an improved alternative assay that could be implemented on a large scale (Weller *et al.*, 2000).

Likewise, food contamination by mycotoxins is of great concern, since many have been found to be carcinogenic and they are not easily removed during food processing. However, since toxin abundance does not correlate with fungal contamination, but is linked to the toxinogenic properties of each strain, real-time PCR detection assays targeted at genes involved in toxinogenesis have been developed for trichotecene-producing *Fusarium* and aflatoxin-producing *Aspergillus* species (Mayer *et al.*, 2003; Schnerr *et al.*, 2001). Petit *et al.* (2003) recently implemented a refinement of this technique based on the quantification of the *nor1* mRNA, which directly addresses aflatoxin biosynthesis in infected wheat. As many countries are becoming more and more concerned about food safety, the market for such applications is growing rapidly.

Real-time PCR application in fundamental studies is still lagging behind, and only a few real-time PCR-based pathogenicity assays have been reported in this field (van Wees *et al.*, 2003; Hiriart *et al.*, 2003). Most of the currently used resistance tests rely on visual assessment of

the symptoms and spore or colony counting. However, Brouwer *et al.* (2003) recently implemented real-time PCR tests to quantify a number of pathogens on *Arabidopsis* and demonstrated that they are a very interesting alternative to classical tests. More details about pathogen detection assays can be found in Schaad and Frederick (2002), McCartney *et al.* (2003), and Schaad *et al.* (2003).

### *Contamination of processed food by foreign DNA*

A requirement for methods capable of accurately quantifying food contaminants has emerged, due to the introduction of stringent food safety regulations. Since most of them enforce a tolerable level of contamination, accurate quantification is of crucial importance for agribusiness companies. Indeed, on the one hand, they have to comply with those maximal authorized levels, while on the other hand, the rejection of batches falsely labelled as contaminated would lead to unnecessary costs. In this context, real-time quantitative PCR is becoming the technique of choice for assessing food contamination or adulteration. Compared with ELISA, PCR assays are easier to develop since they do not require raising specific antibodies. They provide a higher sensitivity and are better suited for the detection of unwanted food ingredients in highly processed food, notably because DNA is more thermo-stable than proteins. For example, the absence of gluten in baby food can be controlled by an amplification test of cereal genes by real-time PCR (Sandberg *et al.*, 2003). Likewise, real-time quantitative PCR has been shown to be an ideal tool for assessing common wheat (*Triticum aestivum*) adulteration in durum wheat pasta (Alary *et al.*, 2002; Terzi *et al.*, 2003), as Spanish, Italian, and French regulations enforce a 3% maximal level of common wheat contamination in pasta and semolina.

Another field of applications was born with the launching of genetically modified organisms on the market. Indeed, the European Community Council recently enacted a new regulation (EC Regulation no. 1829/2003) enforcing the compulsory labelling of food containing more than 0.9% GMOs. However, GMO detection is not trivial and current assays present a number of worrying limitations, which have been nicely reviewed by Ahmed (2002) and Auer (2003). Briefly, protocols aimed at detecting transgenic DNA contamination were developed mainly for transgenic soybean (Berdal and Holst-Jensen, 2001; Hird *et al.*, 2003) and maize (Brodmann *et al.*, 2002; Vaitilingom *et al.*, 1999), and more rarely for other species such as rapeseed (Zeitler *et al.*, 2002). A major limitation of PCR-based detection assays is that a new set of oligonucleotides has to be designed for each transgenic line, except when they are targeted to common DNA regions such as the CaMV35S promoter (Hohne *et al.*, 2002). Alternatively, event-specific protocols have been developed for unique lines, such as Starlink (Windels *et al.*, 2003) and Bt11

(Ronning *et al.*, 2003) transgenic maize. In order to make the detection specific for only one given and identifiable event, scientists have cloned the borders separating the transgene from the rest of the host genome, and used them as specific markers of the given event (Hernández *et al.*, 2003b; Taverniers *et al.*, 2001).

### Genetics of transgenes

Transgenic plants are easily amenable to genetic analyses when the transgene is inserted as a single copy within the host genome. Ingham *et al.* (2001) showed that duplex real-time PCR can be used to determine transgene copy number in transformed plants. They found a strict correlation between their results and Southern blot analyses, except for two lines (out of 37) in which the discrepancy could be ascribed to multiple insertions at a single locus. Overall, they demonstrated that real-time PCR provided a fast and robust method for this application, which could easily be automated and applied to a large number of samples. In lines containing a single insertion, multiplexed PCR could even discriminate between homozygous and hemizygous plants. If previously genotyped lines are not available to calibrate the assay, German *et al.* (2003) proposed using hemizygous T<sub>0</sub> plants as a standard. In this way, homozygous plants can be selected before they produce seeds, thereby avoiding laborious segregation analyses.

## Quantification of specific transcripts

### Integrated expression analyses

Integrative developmental biology requires the parallel analyses of genes involved in the same physiological process. The study of their expression profile may be useful in understanding the cellular function of the encoded proteins. Indeed, in yeast, groups of genes displaying the same expression profile are enriched in genes encoding proteins interacting physically (Ge *et al.*, 2001). However, comparisons between numerous gene expression profiles require a sensitive and reliable technique avoiding errors inherent in independent RNA preparations. Because of its sensitivity, real-time PCR is able to meet this technical requirement, as the expression of numerous genes may be tested on the same RNA preparation. Taking advantage of this, Shimada *et al.* (2003) used real-time PCR to investigate the expression of eight genes involved in brassinosteroid (BR) biosynthesis, degradation or signal transduction. They were able to correlate the expression of the BR synthesizing enzymes with the presence of BR *in planta*, thus providing a more complete view of BR metabolism at the whole plant level. Likewise, Anterola *et al.* (2002) surveyed the expression of genes involved in phenylpropanoid biosynthesis in pine, following phenylalanine feeding. Their study provided very interesting insights about the coordinate regulation of seven

key genes (among them phenylalanine-ammonia-lyase and coumarate-4-hydroxylase) involved in lignin precursors metabolism.

### Gene families

Multigenic families are a distinctive feature of plant genomes, as opposed to animals. For instance, 66% of *A. thaliana* genes belong to families, with half of them containing at least five members (The *Arabidopsis* Genome Initiative, 2000), and information available from ESTs programs on other plant species shows that this particularity is not restricted to *Arabidopsis*. Analysing all the members of a gene family is necessary in order to obtain an accurate view of its overall function. Often, studies describing the expression profile of multigene family members have been performed in *A. thaliana* because of the possibility of an exhaustive survey. However, other plants with a high number of publicly available ESTs, such as wheat, maize, tomato, rice, potato, and the moss *Physcomitrella patens*, could be used for similar analyses. However, the amplification of several highly similar sequences cannot be excluded in those organisms, and data interpretation should be performed very cautiously.

In this context, several requirements have led plant scientists to use real-time RT-PCR methods rather than northern blotting. First, the analysis of more than ten genes by northern blotting is a fairly tedious and repetitive work, as several identical blots have to be prepared (Dong *et al.*, 2003). Instead, Yokoyama and Nishitani (2001) analysed the expression profile of the 33 members of the xyloglucan transglucosylase/hydrolase (XTH) gene family in five organs of *A. thaliana* and in response to four hormones by real-time RT-PCR.

Secondly, genes expressed at a very low level remain difficult to detect by northern blot techniques (Brown *et al.*, 2003; Jakab *et al.*, 2003). In *A. thaliana*, the ARIADNE gene family is composed of 16 genes that have been identified as putative E3-ligases, based on their homology with *Drosophila* and mouse genes involved in the ubiquitin-mediated protein degradation pathway. Their expression was studied by real-time RT-PCR (Mladek *et al.*, 2003). The transcripts of seven genes were undetectable or close to the background level, while the remaining genes were expressed, but with different absolute mRNA levels that varied up to 25-fold. The absolute transcript level remained very low, with fewer than 50 copies per nanogram of total RNA for eight of the nine *AtARI* genes analysed. The sensitivity of real-time RT-PCR showed that, while the expression profile was quasi-constitutive for nearly all of the *AtARI* genes, two of the very low expressed genes displayed a specific expression pattern (Mladek *et al.*, 2003).

Thirdly, closely related genes that are very similar at the sequence level may cross-hybridize during northern blot

procedures (Montrichard *et al.*, 2003), and therefore, it may be difficult to determine the RNA level of a specific member of a gene family. This problem is resolved by the high specificity of real-time RT-PCR guaranteed by the use of at least two specific oligonucleotides (Fig. 2). The GSK3/Shaggy kinase family is composed of ten genes in *A. thaliana* (*AtSK*). Their expression level is rather low (2-fold lower than actin; Charrier *et al.*, 2002), and difficult to study by northern blotting, although possible, as reported by Dornelas *et al.* (1999). However, some of these genes display a very high level of nucleotide sequence identity (up to 98.2%; Charrier *et al.*, 2002), and finding probes specific for each of the ten genes requires them to be designed in the 5' or 3' UTR regions. In order to overcome these difficulties, real-time RT-PCR was used to study the expression profile of the entire gene family, both in several organs of the plant and in response to a number of abiotic stresses. Relative analyses showed that, while most of the *AtSK* genes were expressed constitutively, two of them displayed either some organ specificity or responded to osmotic stress (Charrier *et al.*, 2002).

Knowing the exhaustive expression pattern of a gene family opens up the additional investigation branch of molecular and functional evolution. Indeed, several studies have shown that the expression profile of gene family members was not strictly related to the position of the corresponding proteins within a phylogenetic tree (Mladek *et al.*, 2003; Orsel *et al.*, 2002; Panchuk *et al.*, 2002; Tan *et al.*, 2003; Yokoyama and Nishitani, 2001). This suggested that the regulation of gene expression had undergone different molecular evolutionary mechanisms compared with those influencing protein sequences.

To date, only a limited number of additional studies have described gene family expression profiles using real-time RT-PCR (Balbi and Lomax, 2003; Berger *et al.*, 2002; Kürsteiner *et al.*, 2003; Reintanz *et al.*, 2002; Shimada *et al.*, 2003; Teramoto *et al.*, 2002). Alternative methods to address the transcriptional behaviour of members of large gene families are available, such as microarray hybridization, as illustrated for 142 members of the *Arabidopsis* cytochrome P<sub>450</sub> gene family (Xu *et al.*, 2001). However, due to their specific limitations, an exhaustive picture can only be obtained by the parallel use of those methods.

#### Confirmation of microarray experiments

One of the fastest expanding applications of real-time RT-PCR is the confirmation of data obtained from microarray studies. Indeed, the reliability of microarray experiments may sometimes be questioned. Since plants display a high number of multigene families, cross-hybridization between cDNA representatives of members of gene families on cDNA-based chips may lead to false interpretations. On the other hand, microarray experiments can analyse thousands of genes in one step, whereas real-time PCR is often limited

to far fewer genes. Real-time PCR requires the design of specific oligonucleotides for each gene to be analysed, and because of the limited number of both fluorophores and light spectra detected by real-time PCR machines, this allows the detection of fewer than five genes per multiplex PCR run. However, a maximum of two genes are analysed routinely in the same tube. Therefore, a widely used strategy is to point out a handful of potentially interesting genes with microarray experiments and to confirm those candidates by real-time RT-PCR analysis (Klok *et al.*, 2002). As an illustration, Rider *et al.* (2003) randomly picked 12 candidates among a pool of 185 genes previously identified by Affimetrix<sup>TM</sup> microarray experiments as being up-regulated in the seeds of the *Arabidopsis pickel* mutant compared with wild-type seeds. They confirmed the up-regulation for 10 of those genes by real-time RT-PCR experiments.

While some studies are purely confirmatory, others have used real-time RT-PCR to analyse the expression pattern of the candidate genes further, either to determine fine-tuned kinetics (Goda *et al.*, 2002; Goto and Naito, 2002), or in conditions where the available material was sparse, thereby taking advantage of the technique's sensitivity. The hybridization of cDNA microarrays with RNA of *Arabidopsis* seedlings infected with the incompatible fungus *Alternaria brassicicola* led to the identification of functional groups of genes involved in systemic acquired resistance (Schenk *et al.*, 2003). Then, 23 genes, each representative of one of these groups, were chosen to perform a time-course study in response to *A. brassicicola* infection in local and distal leaf tissues by real-time RT-PCR (Schenk *et al.*, 2003).

Whereas numerous studies have obtained similar results by real-time PCR and microarray experiments, with a linear correlation of up to five orders of magnitude (Maguire *et al.*, 2002), several other articles report that the *n*-fold variation measured by real-time PCR is generally lower (up to 10 times) than that measured using cDNA microarray, where cross-hybridization may occur for highly similar, yet distinct gene sequences (Schenk *et al.*, 2003).

Even with the Affimetrix<sup>TM</sup> chips, which are oligonucleotide-based and therefore more specific than PCR-fragment-based microarray slides, discrepancies between microarray data and real-time PCR data have been noticed. Wang *et al.* (2003) showed that real-time PCR data could display higher induction ratios compared with microarrays, yet conserving a good correlation between the two techniques. However, Hammond *et al.* (2003) noticed that the magnitude and the kinetics of the response of several genes to phosphate starvation differed between the two techniques, including opposite responses for some of them. These conflicting results were explained by the fact that the two experiments were performed with two different biological samples and different oligonucleotides.

In all cases, real-time RT-PCR is considered to be the most reliable technique and discrepancies can most often be ascribed to the normalization or background subtraction

**Table 1.** Summary of the technical qualities of real-time PCR, compared with the other currently used techniques

Number of arrows reflects the level of stringency required from the different parameters. The '+' and the '-' signs indicate the capacity of the technique to satisfy the technical requirement. n.a.: non applicable.

Aim	Rapidity	Sensitivity	Specificity	Quantification	Reliability	Cost
Contamination by foreign DNA						
Technical requirement	→	→→→	→→	→→→	→→→	Relative to laboratory budget
Real-time PCR	+++	+++	+++	+++	+++	-
Southern	-	--	Variable	-	+	+
ELISA tests	+++	--	Variable	+	+	+++
Transgene genetics						
Technical requirement	→→	→	→	→→→	→→→	Relative to laboratory budget
Real-time PCR	+++	+++	+++	+++	+++	-
Genetics	-	n.a.	+	+++	+++	+++
Expression profile						
Technical requirement	→	→→→ (Integrated studies)	→→→ (Gene family)	→→→ (Microarray)	→→→	Relative to laboratory budget
Real-time PCR	+++	+++	+++	+++	+++	-
Oligonucleotide-based microarray	+	+	++	+	+	--
cDNA or PCR product-based microarray	(Variable)	+	-	(Variable)	-	-
Northern blot and dot plot	-	-	+	+	+	+
			(Probe dependent)	(If specific)	(For relative profile)	+

methods used in microarray analysis. Hence, real-time RT-PCR has even been used as a reference to compare different methods of microarray analyses (Puthoff *et al.*, 2003).

The application range of real-time quantitative PCR is much broader than what can possibly be exposed in the framework of this review. For example, in the domain of DNA quantification, Jones *et al.* (2001) nicely exploited real-time PCR to quantify the proportion of restricted DNA corresponding to a transgene, after methylation as a result of post-transcriptional gene silencing. While the rapidity and the reliability of real-time RT-PCR has allowed the description of the expression pattern of numerous individual genes in wild-type plants (Blanco-Portales *et al.*, 2002; Choi *et al.*, 2002; Dambrauskas *et al.*, 2003; Holmberg *et al.*, 2002; Nolan *et al.*, 2003; Shen *et al.*, 2003; Suzuki *et al.*, 2002; Thomas *et al.*, 2003), mutants (Chae *et al.*, 2003) and transgenic plants (Busch *et al.*, 2002; Chang *et al.*, 2003; Itoh *et al.*, 2001), as well as the description of their molecular phenotypes (Balbi and Lomax, 2003; Maruyama-Nakashita *et al.*, 2003), the technique's sensitivity has also been used to calculate gene silencing efficiency (Busch *et al.*, 2002; Lacomme *et al.*, 2003), and to determine the frequency of alternative splicing (Haltermann *et al.*, 2003).

### Limitations and future developments

Despite numerous advantages, real-time PCR has some limitations. Since it is performed on small DNA fragments, real-time quantitative PCR might fail to detect biologically relevant processes like alternative splicing or partial transcript degradation occurring during post-transcriptional

gene silencing events. Performing northern blots in parallel with real-time PCR should help to overcome this difficulty, as suggested by Montrichard *et al.* (2003) in their study of the NADPH/NADP-thioredoxin gene expression. Once specific splicing sites are known, appropriate intron-hybridizing primers can be designed to monitor the accumulation of a specific transcript, as nicely illustrated by Haltermann *et al.* (2003).

For convenience, most real-time PCR analyses are currently performed at the organ level, but further studies may take advantage of the technique's unequalled sensitivity, and address gene expression at the cellular level. For instance, Philippar *et al.* (2003) were able to study the expression pattern of a potassium channel-encoding gene after manually dissecting epidermal, mesophyll, and vascular tissues from maize leaves. Nakazono *et al.* (2003) also nicely demonstrated that laser-capture micro-dissection allowed epidermal cells and vascular tissue to be dissected from maize coleoptiles, obtaining about 40 ng total RNA for each tissue, corresponding to 10 000 cells. However, complementary approaches using reporter genes or *in situ* RNA hybridizations are still useful for addressing gene expression profiles at the cellular level (Berger *et al.*, 2002; Reintanz *et al.*, 2002).

### Conclusions

Real-time quantitative PCR was first developed to meet specific technical requirements, such as a high sensitivity and specificity, which were not easily achieved with other classical techniques. It is now becoming a routine tool, and it is believed that, thanks to its experienced reliability, its

applications will proliferate in the forthcoming years. Thanks to its rapidity, it should even replace some widely used protocols, like Southern blotting for transgenic plant analysis. A comparison of real-time PCR with other laboratory techniques with regard to their most common applications is provided in Table 1. Most evidently, real-time PCR development is still limited by the high costs of the machine and reagents, but hopefully, future will make this technology economically more widely accessible.

## Acknowledgements

We are grateful to Y Henry, A Picaud and M Hodges (IBP, Université Paris-Sud, Orsay, France), F Vedele and M Orsel (INRA, Versailles, France), and FC Küpper (The Scottish Association for Marine Science, Oban, UK) for critical reading of the manuscript and helpful suggestions.

## References

- Ahmed FE. 2002. Detection of genetically modified organisms in foods. *Trends in Biotechnology* **20**, 215–223.
- Alary R, Serin A, Duviau MP, Joudrier P, Gautier MF. 2002. Quantification of common wheat adulteration of durum wheat pasta using real-time quantitative polymerase chain reaction (PCR). *Cereal Chemistry* **79**, 553–558.
- Anterola AM, Jeon JH, Davin LB, Lewis NG. 2002. Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *Journal of Biological Chemistry* **277**, 18272–18280.
- Auer CA. 2003. Tracking genes from seed to supermarket: techniques and trends. *Trends in Plant Science* **8**, 591–597.
- Bago B, Zipfel W, Williams RM, Jun J, Arreola R, Lammers PJ, Pfeffer PE, Shachar Hill Y. 2002. Translocation and utilization of fungal storage lipid in the arbuscular mycorrhizal symbiosis. *Plant Physiology* **128**, 108–124.
- Balbi V, Lomax TL. 2003. Regulation of early tomato fruit development by the *Diageotropica* gene. *Plant Physiology* **131**, 186–197.
- Berdal KG, Holst-Jensen A. 2001. Roundup Ready<sup>®</sup> soybean event-specific real-time quantitative PCR assay and estimation of the practical detection and quantification limits in GMO analyses. *European Food Research and Technology* **213**, 432–438.
- Berger S, Mitchell-Olds T, Stotz HU. 2002. Local and differential control of vegetative storage protein expression in response to herbivore damage in *Arabidopsis thaliana*. *Physiologia Plantarum* **114**, 85–91.
- Blanco-Portales R, Medina-Escobar N, López-Raez JA, González-Reyes JA, Villalba JM, Moyano E, Caballero JL, Muñoz-Blanco J. 2002. Cloning, expression and immunolocalization pattern of a cinnamyl alcohol dehydrogenase gene from strawberry (*Fragaria × ananassa* cv. Chandler). *Journal of Experimental Botany* **53**, 1723–1734.
- Böhm J, Hahn A, Schubert R, Bahnweg G, Adler N, Nechwatal J, Oehlmann R, Osswald W. 1999. Real-time quantitative PCR: DNA determination in isolated spores of the mycorrhizal fungus *Glomus mossae* and monitoring of *Phytophthora infestans* and *Phytophthora citricola* in their respective host plants. *Journal of Phytopathology* **147**, 409–416.
- Bovy A, de Vos R, Kemper M, et al. 2002. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *The Plant Cell* **14**, 2509–2526.
- Brodmann PD, Ilg EC, Berthoud H, Herrmann A. 2002. Real-time quantitative polymerase chain reaction methods for four genetically modified maize varieties and maize DNA content in food. *Journal of AOAC International* **85**, 646–653.
- Brouwer M, Lievens B, Van Hemelrijck W, Van den Ackerveken G, Cammue BP, Thomma BP. 2003. Quantification of disease progression of several microbial pathogens on *Arabidopsis thaliana* using real-time fluorescence PCR. *FEMS Microbiological Letters* **228**, 241–248.
- Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM. 2003. A role for the GCC-box in jasmonate-mediated activation of the PDF1.2 gene of *Arabidopsis*. *Plant Physiology* **132**, 1020–1032.
- Busch M, Seuter A, Hain R. 2002. Functional analysis of the early steps of carotenoid biosynthesis in tobacco. *Plant Physiology* **128**, 439–453.
- Bustin SA. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* **25**, 169–193.
- Bustin SA. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29**, 23–39.
- Chae HS, Faure F, Kieber JJ. 2003. The *eto1*, *eto2*, and *eto3* mutations and cytokinin treatment increase ethylene biosynthesis in *Arabidopsis* by increasing the stability of ACS protein. *The Plant Cell* **15**, 545–559.
- Chang HS, Jones ML, Banowitz GM, Clark DG. 2003. Overproduction of cytokinins in petunia flowers transformed with P-SAG12-IPT delays corolla senescence and decreases sensitivity to ethylene. *Plant Physiology* **132**, 2174–2183.
- Charrier B, Champion A, Henry Y, Kreis M. 2002. Expression profiling of the whole *Arabidopsis* Shaggy-like kinase multigene family by real-time reverse transcriptase-polymerase chain reaction. *Plant Physiology* **130**, 577–590.
- Choi DW, Rodriguez EM, Close TJ. 2002. Barley *Cbf3* gene identification, expression pattern, and map location. *Plant Physiology* **129**, 1781–1787.
- Considine MJ, Daley DO, Whelan J. 2001. The expression of alternative oxidase and uncoupling protein during fruit ripening in mango. *Plant Physiology* **126**, 1619–1629.
- Dambrauskas G, Aves SJ, Bryant JA, Francis D, Rogers HJ. 2003. Genes encoding two essential DNA replication activation proteins, Cdc6 and Mcm3, exhibit very different patterns of expression in the tobacco BY-2 cell cycle. *Journal of Experimental Botany* **54**, 699–706.
- Dong JX, Chen CH, Chen ZX. 2003. Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response. *Plant Molecular Biology* **51**, 21–37.
- Dornelas MC, Wittich P, von Recklinghausen I, van Lammeren A, Kreis M. 1999. Characterization of three novel members of the *Arabidopsis* SHAGGY-related protein kinase (ASK) multigene family. *Plant Molecular Biology* **39**, 137–147.
- Filion M, St-Arnaud M, Jabaji-Hare SH. 2003. Direct quantification of fungal DNA from soil substrate using real-time PCR. *Journal of Microbiological Methods* **53**, 67–76.
- Freeman WM, Walker SJ, Vrana KE. 1999. Quantitative RT-PCR: pitfalls and potential. *Biotechniques* **26**, 112–122, 124–115.
- Ge H, Liu ZH, Church GM, Vidal M. 2001. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genetics* **29**, 482–486.
- German MA, Kandel-Kfir M, Swarzbarg D, Matsevitz T, Granot D. 2003. A rapid method for the analysis of zygosity in transgenic plants. *Plant Science* **164**, 183–187.



- Goda H, Shimada Y, Asami T, Fujioka S, Yoshida S. 2002. Microarray analysis of brassinosteroid-regulated genes in *Arabidopsis*. *Plant Physiology* **130**, 1319–1334.
- Gonçalves MC, Klerks MM, Verbeek M, Vega J, van den Heuvel JF. 2002. The use of molecular beacons combined with NASBA for the sensitive detection of Sugarcane Yellow Leaf Virus. *European Journal of Plant Pathology* **108**, 401–407.
- Goto DB, Naito S. 2002. *AtMRD1* and *AtMRU1*, two novel genes with altered mRNA levels in the methionine over-accumulating *mtol-1* mutant of *Arabidopsis thaliana*. *Plant and Cell Physiology* **43**, 923–931.
- Halterman DA, Wei FS, Wise RP. 2003. Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames. *Plant Physiology* **131**, 558–567.
- Hammond JP, Bennett MJ, Bowen HC, Broadley MR, Eastwood DC, May ST, Rahn C, Swarup R, Woolaway KE, White PJ. 2003. Changes in gene expression in *Arabidopsis* shoots during phosphate starvation and the potential for developing smart plants. *Plant Physiology* **132**, 578–596.
- Harari-Steinberg O, Ohad I, Chamovitz DA. 2001. Dissection of the light signal transduction pathways regulating the two *early light-induced* protein genes in *Arabidopsis*. *Plant Physiology* **127**, 986–997.
- Heid CA, Stevens J, Livak KJ, Williams PM. 1996. Real time quantitative PCR. *Genome Research* **6**, 986–994.
- Hernández M, Ferrando A, Esteve T, Puigdomènech P, Prat S, Pla M. 2003a. Real-time and conventional polymerase chain reaction systems based on the metallo-carboxypeptidase inhibitor gene for specific detection and quantification of potato and tomato in processed food. *Journal of Food Protection* **66**, 1063–1070.
- Hernández M, Pla M, Esteve T, Prat S, Puigdomènech P, Ferrando A. 2003b. A specific real-time quantitative PCR detection system for event MON810 in maize YieldGard® based on the 3'-transgene integration sequence. *Transgenic Research* **12**, 179–189.
- Higuchi R, Fockler C, Dollinger G, Watson R. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/technology* **11**, 1026–1030.
- Hird H, Powell J, Johnson ML, Oehlschlager S. 2003. Determination of percentage of RoundUp Ready® soya in soya flour using real-time polymerase chain reaction: Interlaboratory study. *Journal of AOAC International* **86**, 66–71.
- Hiriart JB, Aro EM, Lehto K. 2003. Dynamics of the VIGS-mediated chimeric silencing of the *Nicotiana benthamiana* ChH gene and of the Tobacco Mosaic Virus Vector. *Molecular Plant-Microbe Interactions* **16**, 99–106.
- Hohne M, Santisi CR, Meyer R. 2002. Real-time multiplex PCR: an accurate method for the detection and quantification of 35S-CaMV promoter in genetically modified maize-containing food. *European Food Research and Technology* **215**, 59–64.
- Holmberg N, Harker M, Gibbard CL, Wallace AD, Clayton JC, Rawlins S, Hellyer A, Safford R. 2002. Sterol C-24 methyltransferase type 1 controls the flux of carbon into sterol biosynthesis in tobacco seed. *Plant Physiology* **130**, 303–311.
- Ingham DJ, Beer S, Money S, Hansen G. 2001. Quantitative real-time PCR assay for determining transgene copy number in transformed plants. *Biotechniques* **31**, 132, 134, 136–140.
- Itoh R, Fujiwara M, Nagata N, Yoshida S. 2001. A chloroplast protein homologous to the eubacterial topological specificity factor MinE plays a role in chloroplast division. *Plant Physiology* **127**, 1644–1655.
- Jakab G, Manrique A, Zimmerli L, Métraux JP, Mauch-Mani B. 2003. Molecular characterization of a novel lipase-like pathogen-inducible gene family of *Arabidopsis*. *Plant Physiology* **132**, 2230–2239.
- Jones L, Ratcliff F, Baulcombe DC. 2001. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Current Biology* **11**, 747–757.
- Klok EJ, Wilson IW, Wilson D, Chapman SC, Ewing RM, Somerville SC, Peacock WJ, Dolferus R, Dennis ES. 2002. Expression profile analysis of the low-oxygen response in *Arabidopsis* root cultures. *The Plant Cell* **14**, 2481–2494.
- Kürsteiner O, Dupuis I, Kuhlemeier C. 2003. The *Pyruvate decarboxylase1* gene of *Arabidopsis* is required during anoxia, but not other environmental stresses. *Plant Physiology* **132**, 968–978.
- Lacomme C, Hrubikova K, Hein I. 2003. Enhancement of virus-induced gene silencing through viral-based production of inverted repeats. *The Plant Journal* **34**, 543–553.
- Maguire TL, Grimmond S, Forrest A, Iturbe Ormaetxe I, Meksem K, Gresshoff P. 2002. Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis. *Journal of Plant Physiology* **159**, 1361–1374.
- Maruyama-Nakashita A, Inoue E, Watanabe-Takahashi A, Yamaya T, Takahashi H. 2003. Transcriptome profiling of sulfur-responsive genes in *Arabidopsis* reveals global effects of sulfur nutrition on multiple metabolic pathways. *Plant Physiology* **132**, 597–605.
- Mayer Z, Bagnara A, Färber P, Geisen R. 2003. Quantification of the copy number of *nor-1*, a gene of the aflatoxin biosynthetic pathway by real-time PCR, and its correlation to the cfu of *Aspergillus flavus* in foods. *International Journal of Food Microbiology* **82**, 143–151.
- McCartney HA, Foster SJ, Fraaije BA, Ward E. 2003. Molecular diagnostics for fungal plant pathogens. *Pest Management Science* **59**, 129–142.
- Mladek C, Guger K, Hauser MT. 2003. Identification and characterization of the ARIADNE gene family in *Arabidopsis*. A group of putative E3 ligases. *Plant Physiology* **131**, 27–40.
- Montrichard F, Renard M, Alkhalfioui F, Duval FD, Macherel D. 2003. Identification and differential expression of two thioredoxin h isoforms in germinating seeds from pea. *Plant Physiology* **132**, 1707–1715.
- Nakazono M, Qiu F, Borsuk LA, Schnable PS. 2003. Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: Identification of genes expressed differentially in epidermal cells or vascular tissues of maize. *The Plant Cell* **15**, 583–596.
- Nolan KE, Irwanto RR, Rose RJ. 2003. Auxin up-regulates *MtSERK1* expression in both *Medicago truncatula* root-forming and embryogenic cultures. *Plant Physiology* **133**, 218–230.
- Orsel M, Krapp A, Daniel-Vedele F. 2002. Analysis of the NRT2 nitrate transporter family in *Arabidopsis*. Structure and gene expression. *Plant Physiology* **129**, 886–896.
- Panchuk II, Volkov RA, Schöffl F. 2002. Heat stress- and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in *Arabidopsis*. *Plant Physiology* **129**, 838–853.
- Petit L, Baraige F, Balois AM, Bertheau Y, Fach P. 2003. Screening of genetically modified organisms and specific detection of Bt176 maize in flours and starches by PCR-enzyme linked immunosorbent assay. *European Food Research and Technology* **217**, 83–89.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45.
- Phillippar K, Büchenschütz K, Abshagen M, Fuchs I, Geiger D, Lacomme B, Hedrich R. 2003. The K<sup>+</sup> channel KZM1 mediates potassium uptake into the phloem and guard cells of the C<sub>4</sub> grass *Zea mays*. *Journal of Biological Chemistry* **278**, 16973–16981.

- Puthoff DP, Nettleton D, Rodermel SR, Baum TJ.** 2003. *Arabidopsis* gene expression changes during cyst nematode parasitism revealed by statistical analyses of microarray expression profiles. *The Plant Journal* **33**, 911–921.
- Reintanz B, Szyroki A, Ivashikina N, Ache P, Godde M, Becker D, Palme K, Hedrich R.** 2002. AtKC1, a silent *Arabidopsis* potassium channel alpha-subunit modulates root hair K<sup>+</sup> influx. *Proceedings of the National Academy of Sciences, USA* **99**, 4079–4084.
- Rider SD, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J.** 2003. Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis*. *The Plant Journal* **35**, 33–43.
- Ronning SB, Vaitilingom M, Berdal KG, Holst-Jensen A.** 2003. Event specific real-time quantitative PCR for genetically modified Bt11 maize (*Zea mays*). *European Food Research and Technology* **216**, 347–354.
- Sandberg M, Lundberg L, Ferm M, Yman IM.** 2003. Real-time PCR for the detection and discrimination of cereal contamination in gluten free foods. *European Food Research and Technology* **217**, 344–349.
- Schaad NW, Frederick RD.** 2002. Real-time PCR and its application for rapid plant disease diagnostics. *Canadian Journal of Plant Pathology* **24**, 250–258.
- Schaad NW, Frederick RD, Shaw J, Schneider WL, Hickson R, Petrillo MD, Luster DG.** 2003. Advances in molecular-based diagnostics in meeting crop biosecurity and phytosanitary issues. *Annual Review of Phytopathology* **41**, 305–324.
- Schena L, Nigro F, Ippolito A.** 2002. Identification and detection of *Rosellinia necatrix* by conventional and real-time Scorpion-PCR. *European Journal of Plant Pathology* **108**, 355–366.
- Schenk PM, Kazan K, Manners JM, Anderson JP, Simpson RS, Wilson IW, Somerville SC, Maclean DJ.** 2003. Systemic gene expression in *Arabidopsis* during an incompatible interaction with *Alternaria brassicicola*. *Plant Physiology* **132**, 999–1010.
- Schnerr H, Niessen L, Vogel RF.** 2001. Real time detection of the *tri5* gene in *Fusarium* species by LightCycler<sup>TM</sup>-PCR using SYBR<sup>®</sup>-Green I for continuous fluorescence monitoring. *International Journal of Food Microbiology* **71**, 53–61.
- Shen GA, Pang YZ, Lin CF, Wei C, Qian XY, Jiang LZ, Du XL, Li KG, Attia K, Yang JS.** 2003. Cloning and characterization of a novel Hsp100/Clp gene (*OsClpD*) from *Oryza sativa*. *DNA Sequence* **14**, 285–293.
- Shimada Y, Goda H, Nakamura A, Takatsuto S, Fujioka S, Yoshida S.** 2003. Organ-specific expression of brassinosteroid-biosynthetic genes and distribution of endogenous brassinosteroids in *Arabidopsis*. *Plant Physiology* **131**, 287–297.
- Suzuki H, Nakayama T, Yonekura-Sakakibara K, Fukui Y, Nakamura N, Yamaguchi M, Tanaka Y, Kusumi T, Nishino T.** 2002. cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A:anthocyanidin 3-O-glucoside-6"-O-malonyltransferase from dahlia flowers. *Plant Physiology* **130**, 2142–2151.
- Tan BC, Joseph LM, Deng WT, Liu LJ, Li QB, Cline K, McCarty DR.** 2003. Molecular characterization of the *Arabidopsis* 9-cis epoxycarotenoid dioxygenase gene family. *The Plant Journal* **35**, 44–56.
- Taverniers I, Windels P, Van Bockstaele E, De Loose M.** 2001. Use of cloned DNA fragments for event-specific quantification of genetically modified organisms in pure and mixed food products. *European Food Research and Technology* **213**, 417–424.
- Teramoto H, Nakamori A, Minagawa J, Ono T.** 2002. Light-intensity-dependent expression of Lhc gene family encoding light-harvesting chlorophyll-*a/b* proteins of photosystem II in *Chlamydomonas reinhardtii*. *Plant Physiology* **130**, 325–333.
- Terzi V, Malnati M, Barbanera M, Stanca AM, Faccioli P.** 2003. Development of analytical systems based on real-time PCR for *Triticum* species-specific detection and quantitation of bread wheat contamination in semolina and pasta. *Journal of Cereal Science* **38**, 87–94.
- The Arabidopsis Genome Initiative.** 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Thomas C, Meyer D, Wolff M, Hember C, Alioua M, Steinmetz A.** 2003. Molecular characterization and spatial expression of the sunflower ABP1 gene. *Plant Molecular Biology* **52**, 1025–1036.
- Vaitilingom M, Pijenburg H, Gendre F, Brignon P.** 1999. Real-time quantitative PCR detection of genetically modified Maximizer maize and RoundUp Ready soybean in some representative foods. *Journal of Agricultural and Food Chemistry* **47**, 5261–5266.
- van Wees SCM, Chang HS, Zhu T, Glazebrook J.** 2003. Characterization of the early response of *Arabidopsis* to *Alternaria brassicicola* infection using expression profiling. *Plant Physiology* **132**, 606–617.
- Wang RC, Okamoto M, Xing XJ, Crawford NM.** 2003. Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. *Plant Physiology* **132**, 556–567.
- Weller SA, Elphinstone JG, Smith NC, Boonham N, Stead DE.** 2000. Detection of *Ralstonia solanacearum* strains with a quantitative, multiplex, real-time, fluorogenic PCR (TaqMan<sup>®</sup>) assay. *Applied and Environmental Microbiology* **66**, 2853–2858.
- Windels P, Bertrand S, Depicker A, Moens W, Van Bockstaele E, Loose M.** 2003. Qualitative and event-specific PCR real-time detection methods for StarLink maize. *European Food Research and Technology* **216**, 259–263.
- Winton LM, Manter DK, Stone JK, Hansen EM.** 2003. Comparison of biochemical, molecular, and visual methods to quantify *Phaeocryptopus gaeumannii* in Douglas-Fir foliage. *Phytopathology* **93**, 121–126.
- Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP.** 1997a. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22**, 130–131, 134–138.
- Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ.** 1997b. The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22**, 176–181.
- Wurmbach E, Yuen T, Sealfon SC.** 2003. Focused microarray analysis. *Methods* **31**, 306–316.
- Xu WY, Bak S, Decker A, Paquette SM, Feyereisen R, Galbraith DW.** 2001. Microarray-based analysis of gene expression in very large gene families: the cytochrome P<sub>450</sub> gene superfamily of *Arabidopsis thaliana*. *Gene* **272**, 61–74.
- Yokoyama R, Nishitani K.** 2001. A comprehensive expression analysis of all members of a gene family encoding cell-wall enzymes allowed us to predict *cis*-regulatory regions involved in cell-wall construction in specific organs of *Arabidopsis*. *Plant and Cell Physiology* **42**, 1025–1033.
- Zeitler R, Pietsch K, Waiblinger HU.** 2002. Validation of real-time PCR methods for the quantification of transgenic contaminations in rape seed. *European Food Research and Technology* **214**, 346–351.

```

Sub Modifie1_TraitementFichierCSV()
'
' Modifie1_TraitementFichierCSV Macro
'
' MODIFICATIONS:
'   - repertoire et sous-repertoires des fichiers à traiter
'   - seuil fixe (0.5)
'
' Macro créée le 01/10/2001 par Annaïck
' Racourci clavier : Ctrl+t
'
' Cette macro permet de transformer un fichier .csv en fichier .xls
'
' Un fichier .csv contient des lignes de valeurs séparées par des virgules
' Il contient des résultats de PCR quantitative
' Il a été généré par le logiciel GeneAmp5700SDS (Applied Biosystems)
'
' Le fichier .xls final est composé de 2 feuilles :
' Une contenant les résultats bruts mis en page, appelée "Data"
' l'autre contenant des résultats analysés, appelée "Bilan"
'
' CONSTANTES
'   repertoire = "G:\Annaïck\NouvellesAnalyses\"
'   sousrepCSV = "NouveauxResultatsCSV\"
'   sousrepExcel = "NouveauxFichiersExcel\"
'   seuil = "0,5"
'
' VARIABLES
'   Declaration
'   Dim chemin As String
'   Dim cledutri As String
'   Dim coupleAmorces As String
'   Dim coupleAvant As String
'   Dim debutplagegrise As String
'   Dim debutplageefface As String
'   Dim echantillon1 As String
'   Dim echantillon2 As String
'   Dim finplagegrise As String
'   Dim nom As String
'   Dim nomGene As String
'   Dim pente As String
'   Dim pivot As Range
'   Dim efficacite As Integer
'   Dim feuille As Integer
'   Dim fin As Integer
'   Dim IndiceFeuille As Integer
'   Dim ligne As Integer
'   Dim nbcase As Integer
'   Dim nbgenes As Integer
'   Dim nbl As Integer
'   Dim nblignes As Integer
'   Dim nbLignesTab As Integer
'   Dim minim As Double
'   Dim maxim As Double
'
'   Initialisation
'   pente = "3,32"

```

```

' *****
' *   Création du fichier temporaire contenant les données à traiter *
' *****

' ouverture du fichier .csv
chemin = InputBox("nom du fichier csv a traiter ?" & Chr(10) & _
    "chemin par défaut :" & Chr(10) & repertoire & sousrepCSV, , _
    "results02-05-27")
Workbooks.Open Filename:=(repertoire & sousrepCSV & chemin & ".csv")

' sauvegarde sous Excel dans le fichier temporaire.xls
ChDir repertoire
ActiveWorkbook.SaveAs Filename:="temporaire.xls", FileFormat:= _
    xlNormal, Password:="", WriteResPassword:="", _
    ReadOnlyRecommended:=False, CreateBackup:=False

' *****
' *                               Mise en forme générale                               *
' *****

' suppression de la colonne A (références des puits de la plaque PCR)
Columns("A:A").Select
Selection.Delete Shift:=xlToLeft

' tranformation des titres anglais => francais
Range("B1").FormulaR1C1 = "Nom"
Range("C1").FormulaR1C1 = "Amorces"
Range("F1").FormulaR1C1 = "Qte"
Range("G1").FormulaR1C1 = "Moy Qte"
Range("H1").FormulaR1C1 = "StdDev Qte"

' 1ere ligne = titres => texte en gras, centré et encadré
Range("A1:H1").Select
Selection.HorizontalAlignment = xlCenter
Selection.Font.Bold = True
Selection.Borders(xlEdgeTop).LineStyle = xlContinuous
Selection.Borders(xlEdgeBottom).LineStyle = xlContinuous

'alignement centré du texte dans la colonne "C"
Range("C:C").Select
Selection.HorizontalAlignment = xlCenter

' modification du format des cellules "Ct" et "Qte"
Range("D:E").Select
Selection.NumberFormat = "0.00"
Range("F:H").Select
Selection.NumberFormat = "0"
'remplacement des points par des virgules dans la colonne B
'(incompatibilité excel anglais/français)
Columns("B:B").Select
Selection.Replace What:=".", Replacement:=",", LookAt:=xlPart, _
    SearchOrder:=xlByRows, MatchCase:=False

'tri des données par gène, puis par type de matrice,
'puis par échantillon
Cells.Select
Selection.Sort Key1:=Range("C2"), Order1:=xlAscending, _
    Key2:=Range("A2"), Order2:=xlAscending, Key3:=Range("B2"), _
    Order3:=xlAscending, Header:=xlYes, OrderCustom:=1, _
    MatchCase:=False, Orientation:=xlTopToBottom

```

```

' *****
' *                               Séparation des données gène par gène                               *
' *****

' Sont créées autant de feuilles qu'il n'y a de gènes
' Chaque feuille contient les données d'un seul gène

nbgenes = InputBox("nombre de genes differents ? ", , 1)
If nbgenes > 1 Then
    For feuille = 2 To nbgenes
        ' Recherche des données du gène
        Range("C2").Select
        coupleAvant = ActiveCell.Value
        While ActiveCell.Value = coupleAvant
            ActiveCell.Offset(1, 0).Activate
        Wend
        Set pivot = ActiveCell
        coupleAvant = ActiveCell.Value
        While ActiveCell.Value = coupleAvant
            ActiveCell.Offset(1, 0).Activate
        Wend
        ' Selection des données
        Range(pivot, ActiveCell.Offset(-1, 0)).EntireRow.Select
        ' Coupé-collé des données dans une nouvelle feuille
        Selection.Cut
        Sheets.Add
        ActiveSheet.Paste
        Sheets(2).Select
        Selection.Delete Shift:=xlUp
        Rows("1:1").Select
        Selection.Copy
        Sheets(1).Select
        Range("A1").Select
        Selection.Insert Shift:=xlDown
        ' Déplacement de la nouvelle feuille en dernière position
        Sheets(1).Move After:=Sheets(feuille)
        ' Retour aux données initiales (feuille 1)
        Sheets(1).Activate
    Next feuille
End If

```

```

' *****
' *                               Traitements des données gène par gène *
' *****

For feuille = 1 To nbgenes
' Pour chaque gène
' 1) Mise en forme et analyse des données brutes (feuille Data)
' 2) les données d'expression sont recopiées sur la feuille Bilan,
' mises en forme, analysées et résumées dans un tableau synthétique

' *****      Mise en forme      feuille Data      *****

' changement du nom de la feuille en AtRHnData
' (avec n = numéro du gène)
Sheets((feuille - 1) * 2 + 1).Activate
nomGene = InputBox("nom du gene ? ", , "AtRH" & feuille)
ActiveSheet.Name = nomGene & "Data"
' remplacement du code des amorces "PR*" de la colonne C
' par l'entree utilisateur
Range("C2").Select
coupleAmorces = InputBox("Couple d'amorces ?", , "629/630")
While Not (ActiveCell.Value = "")
    ActiveCell.Value = coupleAmorces
    ActiveCell.Offset(1, 0).Activate
Wend

' Mise en page des valeurs de la gamme (STND) :
' recherche des plages de cellules concernées
Range("A2").Select
While Not (ActiveCell.Value = "STND")
    ActiveCell.Offset(1, 0).Activate
Wend
debutplagegrise = ActiveCell.Address
debutplageefface = ActiveCell.Offset(0, 6).Address
While (ActiveCell.Value = "STND")
    ActiveCell.Offset(1, 0).Activate
Wend
finplagegrise = ActiveCell.Offset(-1, 5).Address
' suppression des valeurs des colonnes "moyenne" et "ecartype",
Range(debutplageefface & ":" & ActiveCell.Offset(-1, _
7).Address).Select
Selection.ClearContents
' fond des cellules en gris
Range(debutplagegrise & ":" & finplagegrise).Select
Selection.Interior.ColorIndex = 15
' tri par la qté d'ADN croissante (colonne F)
cledutri = Right(debutplagegrise, 1)
Selection.Sort Key1:=Range("F" & cledutri)

' Seuil, pente et efficacité (saisie utilisateur + calcul)
Range("H3").Select
ActiveCell.FormulaR1C1 = "seuil = " & seuil

Range("H4").Select
pente = InputBox("pente ?", , pente)
ActiveCell.FormulaR1C1 = "pente = - " & pente

Range("H5").Select
efficacite = ((10 ^ (1 / pente)) - 1) * 100
ActiveCell.FormulaR1C1 = "E = " & Format(efficacite, 0) & "%"

```

```
'encadrement des cellules H3:H5 et texte centré
Range("H3:H5").Select
Selection.Borders(xlEdgeLeft).LineStyle = xlContinuous
Selection.Borders(xlEdgeTop).LineStyle = xlContinuous
Selection.Borders(xlEdgeBottom).LineStyle = xlContinuous
Selection.Borders(xlEdgeRight).LineStyle = xlContinuous
Selection.HorizontalAlignment = xlCenter

' ligne de fin du tableau
nbLignesTab = 0
Range("A1").Select
While Not (ActiveCell.Value = "")
    nbLignesTab = nbLignesTab + 1
    ActiveCell.Offset(1, 0).Activate
Wend
Range("A" & nbLignesTab & ":H" & nbLignesTab).Select
Selection.Borders(xlEdgeBottom).LineStyle = xlContinuous
nbLignesTab = 0
```

```
' *****
                                CALCULS   feuille Data   *****
'
' Recherche des Ct max et min des STND
Range("A2").Select
While Not (ActiveCell.Value = "STND")
    ActiveCell.Offset(1, 0).Activate
Wend
minim = ActiveCell.Offset(0, 3).Value
maxim = ActiveCell.Offset(0, 3).Value
While (ActiveCell.Value = "STND")
    If ActiveCell.Offset(0, 3).Value < minim Then
        minim = ActiveCell.Offset(0, 3).Value
    End If
    If ActiveCell.Offset(0, 3).Value > maxim Then
        maxim = ActiveCell.Offset(0, 3).Value
    End If
    ActiveCell.Offset(1, 0).Activate
Wend

' mise en italique des UNKN hors gamme
nom = ""
While Not (ActiveCell.Value = "")
    If ActiveCell.Offset(0, 3).Value < minim
    Or ActiveCell.Offset(0, 3).Value > maxim Then
        nom = ActiveCell.Offset(0, 1).Value
        While ActiveCell.Offset(-1, 1).Value = nom
            ActiveCell.Offset(-1, 0).Activate
        Wend
        While ActiveCell.Offset(0, 1).Value = nom
            ActiveCell.EntireRow.Font.Italic = True
            ActiveCell.Offset(1, 0).Activate
        Wend
    End If
    ActiveCell.Offset(1, 0).Activate
Wend

' ajustement automatique de la largeur des colonnes
Cells.Select
Selection.Columns.AutoFit
```



```
' *****  
      Mise en forme   feuille Bilan   *****  
  
      ' copie des données sur une nouvelle feuille  
      Selection.Copy  
      Sheets.Add  
      ActiveSheet.Paste  
      ActiveSheet.Name = nomGene & "Bilan"  
      ActiveSheet.Move After:=Sheets(2 * feuille)  
      Range("D:E").Select  
      Selection.ClearContents  
      Columns("E:E").Delete Shift:=xlToLeft  
      Range("D1").FormulaR1C1 = "Qte (ng)"  
      Range("E1").FormulaR1C1 = "Eq.g"  
      Columns("F:F").Select  
      Selection.Insert Shift:=xlToRight  
      Range("F1").FormulaR1C1 = "Eq.g/ng"  
      Range("G1").FormulaR1C1 = "Moy Eq.g/ng"  
      Range("H1").FormulaR1C1 = "StdDev"  
  
      ' suppression des lignes NTC et STND  
      Range("A2").Select  
      While ActiveCell.Value = "NTC" Or ActiveCell.Value = "STND"  
          ActiveCell.Offset(1, 0).Activate  
      Wend  
      Range("A2:" & ActiveCell.Offset(-1, 0).Address).EntireRow.Select  
      Selection.Delete Shift:=xlUp  
      Columns("A:A").Delete Shift:=xlToLeft  
  
      ' suppression des UNKN hors gamme  
      Range("A2").Select  
      While Not ActiveCell.Value = ""  
          If ActiveCell.Font.Italic = True Then  
              ActiveCell.EntireRow.Delete  
          Else  
              ActiveCell.Offset(1, 0).Activate  
          End If  
      Wend
```

```

' *****          CALCULS   feuille Bilan          *****
' le tableau contient-il des données ?
Range("A2").Select
fin = 1
While Not ActiveCell.Value = ""
    fin = fin + 1
    ActiveCell.Offset(1, 0).Activate
Wend

' Si le tableau n'est pas vide
If fin > 1 Then
' Récupération de la quantité d'ADNc déposée et
' calcul du nombre de copies/ng
Range("C2").Select
ActiveCell.FormulaR1C1 = "=VALUE(MID(RC[-2],SEARCH("-", _
    RC[-2],1)+1,LEN(RC[-2])))"
Range("E2").Select
ActiveCell.FormulaR1C1 = "=RC[-1]/RC[-2]"
Selection.Copy
Range("F2").Select
Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
    SkipBlanks:=False, Transpose:=False

'Copie incrémentée des formules si nécessaire
If fin > 2 Then ' début du tableau des données d'expression
Range("C2").Select
Selection.AutoFill Destination:=Range("C2:C" & fin), _
    Type:=xlFillDefault
Range("E2").Select
Selection.AutoFill Destination:=Range("E2:E" & fin), _
    Type:=xlFillDefault

' moyennes et écart-types
Range("F2:G" & fin).Select
Selection.ClearContents
Range("A2").Activate
nom = ""
While Not ActiveCell.Value = ""
    nom = ActiveCell.Value
    nblignes = 0
    While ActiveCell.Value = nom
        nblignes = nblignes + 1
        ActiveCell.Offset(1, 0).Activate
    Wend
    ActiveCell.Offset(-nblignes, 5).FormulaR1C1 =
        "=AVERAGE(RC[-1]:R[" & nblignes - 1 & "]C[-1])"
    ActiveCell.Offset(-nblignes, 6).FormulaR1C1 =
        "=STDEV(RC[-2]:R[" & nblignes - 1 & "]C[-2])"
Wend
Range("D2:G" & fin).Select
Selection.NumberFormat = "0"

```

```
' mise en gris des cellules un échantillon sur deux
Range("A2").Select
echantillon1 = ""
echantillon2 = ""

While Not ActiveCell.Value = ""
    echantillon1 = Left(ActiveCell.Value, InStr _
        (ActiveCell.Value, "-"))
    While Left(ActiveCell.Value, InStr(ActiveCell.Value, _
        "- ")) = echantillon1
        ActiveCell.Offset(1, 0).Activate
    Wend
    If ActiveCell.Value <> "" Then
        echantillon2 = Left(ActiveCell.Value, InStr _
            (ActiveCell.Value, "-"))
        nbl = 0
        While Left(ActiveCell.Value, InStr _
            (ActiveCell.Value, "-")) = echantillon2
            nbl = nbl + 1
            ActiveCell.Offset(1, 0).Activate
            debut = ActiveCell.Row - nbl
        Wend
        Range("A" & debut & ":G" & debut + nbl - 1).Select
        Selection.Interior.ColorIndex = 15
        Range("A" & debut + nbl).Activate
        Else: ActiveCell.Offset(1, 0).Activate
    End If
Wend

'ligne fin du tableau
Range("A" & fin & ":G" & fin).Select
Selection.Borders(xlEdgeBottom).LineStyle = xlContinuous
End If
'fin de mise en page du tableau de données d'expression
```

```

*****      Tableau résumé des résultats      *****
  ' copie des résultats
Range("A1").CurrentRegion.Select
Selection.Copy
Range("I1").Select

  ' collage spécial : "uniquement les valeurs"

Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
  SkipBlanks:=False, Transpose:=False
Columns("J:M").Select
Application.CutCopyMode = False
Selection.Delete Shift:=xlToLeft

  'suppression des lignes vides :
Range("I2").Select
While Not ActiveCell.Value = ""
  If ActiveCell.Offset(0, 1).Value = "" Then
    ligne = ActiveCell.Row
    Range("I" & ligne & ":K" & ligne).Select
    Selection.Delete Shift:=xlUp
  Else
    ActiveCell.Offset(1, 0).Activate
  End If
Wend
Range("I:K").NumberFormat = "0"

'-----      mise en forme du tableau synthétique      -----
  ' titres en gras
Rows("1:1").Select
Selection.Font.Bold = True

  'encadrement
Range("I1").Select
nbcase = 0
While Not ActiveCell.Value = ""
  nbcase = nbcase + 1
  ActiveCell.Offset(1, 0).Activate
Wend
Range("I1:K" & nbcase).Select
Selection.Borders(xlEdgeLeft).LineStyle = xlContinuous
Selection.Borders(xlEdgeTop).LineStyle = xlContinuous
Selection.Borders(xlEdgeBottom).LineStyle = xlContinuous
Selection.Borders(xlEdgeRight).LineStyle = xlContinuous

Range("I1:K1").Select
Selection.Borders(xlEdgeBottom).LineStyle = xlContinuous

Range("I1:I" & nbcase).Select
Selection.Borders(xlEdgeRight).LineStyle = xlContinuous
End If

' fin des calculs et de la mise en forme du tableau synthétique

' Ajustement automatique de la largeur de toutes les colonnes
Cells.Select
Selection.Columns.AutoFit
Range("A1").Select
Next feuille
' Boucle pour traiter les données du gène suivant

```

```
' *****  
' *                               Réinitialisations finales                               *  
' *****  
  
' Selection de la case A1 pour toutes les feuilles  
For IndiceFeuille = 1 To nbgenes  
    Sheets(IndiceFeuille).Select  
        Range("A1").Select  
Next IndiceFeuille  
  
' *****  
' *                               Sauvegarde                               *  
' *****  
  
ActiveWorkbook.SaveAs Filename:=(repertoire & sousrepExcel & chemin)  
  
End Sub
```

RESUME

L'évolution de la régulation de l'expression des gènes participe à l'évolution de leur fonction. L'existence de profils d'expression spécifiques évitant la redondance fonctionnelle expliquerait le maintien des gènes dupliqués dans les génomes. La PCR quantitative en temps réel se révèle adaptée à l'étude de l'expression des familles de gènes présentant des niveaux très variés et des séquences proches. Le génome modèle d'*Arabidopsis thaliana* contient une famille d'ARN hélicases à boîte DEAD (RH) de 58 membres, *i.e.* environ deux fois plus que n'en comptent les génomes d'animaux ou celui de la levure. L'expression transcriptionnelle de 20 *AtRH* a été obtenue par RT-PCR quantitative dans neuf organes différents. Dans cette famille de gènes « de ménage », deux *AtRH* présentent des profils spécifiques alors que les 18 autres *AtRH* présentent le même profil spatial, mais des niveaux d'expression transcriptionnelle très différents. L'élément régulateur principal du niveau transcriptionnel est la présence simultanée d'une boîte TATA caractéristique et d'un intron en 5' UTR. Dès lors que la boîte TATA est présente, il y a une corrélation positive significative entre la taille de l'intron en 5' UTR et le niveau d'expression. Notre travail sur l'expression des *AtRH* permet d'introduire un scénario sur la dynamique évolutive des gènes dupliqués qui forment une même branche terminale d'un arbre phylogénétique et dont le niveau d'expression diffère fréquemment. Après la duplication d'un gène fortement transcrit, l'altération de l'activité transcriptionnelle des copies se produirait par des événements successifs de suppression de la boîte TATA et/ou de l'intron en 5' UTR.

TITLE

Expression and evolution of the DEAD-box RNA helicase gene family in *Arabidopsis thaliana*

ABSTRACT

The evolution of the regulation of gene expression probably played an important part in gene subfunctionalisation and neofunctionalisation. The existence of gene specific transcript profiles suggests functional diversification, hence loss of functional redundancy, and could explain the maintenance of numerous paralogues in a genome. Real-time quantitative RT-PCR techniques proved to be suitable for the expression studies of gene families with highly variable expression levels and related sequences. The model genome of *Arabidopsis thaliana* contains a DEAD box RNA helicase family (RH) of 58 members, *i.e.* almost twice as many as in the animal or yeast genomes. Transcript profiling using real-time quantitative PCR has been obtained for 20 *AtRHs* from 9 different organs. Among the members of this housekeeping gene family, two *AtRHs* exhibited plant specific transcript profiles while the other 18 *AtRHs* had similar profiles, but very different levels of transcription. The main regulatory element of the transcript level is the simultaneous presence of a TATA-box and an intron in the 5' UTR. There is a positive and highly significant correlation between the size of the 5' UTR intron and the transcription level, as long as a characteristic TATA-box is present. Our work on the expression of the *AtRH* genes suggests a scenario for the evolution of duplicated genes in the same terminal branch of a phylogenetic tree and that often present different expression levels. After the duplication of a highly transcribed ancestral *AtRH*, an alteration of the transcriptional activity of the divergent duplicates occurs through successive events of suppression of the TATA-box and/or the 5' UTR intron.

DISCIPLINE / SPECIALITE

Biologie / Biologie Moléculaire Végétale

MOTS-CLES

ARN hélicases à boîte DEAD, PCR quantitative en temps réel, éléments *cis*-régulateurs, profils transcriptionnels

INTITULE ET ADRESSE DU LABORATOIRE

Laboratoire Génome, Evolution et Développement de la Plante  
 Institut de Biotechnologie des Plantes, UMR CNRS 8618  
 Bâtiment 630 - Université Paris Sud  
 91405 ORSAY cedex