



**HAL**  
open science

# Etude de la voie de biosynthese des monolignols chez brachypodium distachyon

Madeleine Bouvier d'Yvoire

► **To cite this version:**

Madeleine Bouvier d'Yvoire. Etude de la voie de biosynthese des monolignols chez brachypodium distachyon. Sciences agricoles. Université Paris Sud - Paris XI, 2011. Français. NNT : 2011PA112347 . tel-00767480

**HAL Id: tel-00767480**

**<https://theses.hal.science/tel-00767480>**

Submitted on 20 Dec 2012

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

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

# THESE

Pour l'obtention du Grade de  
DOCTEUR DE L'UNIVERSITE DE PARIS 11, ORSAY

Discipline : Sciences de la Vie  
École doctorale Sciences du Végétal : du gène à l'écosystème (ED145)

Présentée par

Madeleine BOUVIER d'YVOIRE

\*\*\*\*\*

## Identification de gènes impliqués dans la voie de biosynthèse des lignines chez *Brachypodium distachyon*

\*\*\*\*\*

Soutenue le 19 décembre 2011 devant la commission d'examen :

Michel DRON	Professeur Université Paris XI, Orsay	Président
Jacqueline GRIMA-PETTENATI	Directeur de recherche CNRS	Rapporteur
Simon HAWKINS	Professeur USTL, Lille	Rapporteur
David POT	Chargé de recherche, CIRAD	Examineur
Claire HALPIN	Professeur College of Life Science, Dundee	Examineur
Lise JOUANIN	Directeur de recherche CNRS	Dir. De thèse
Catherine LAPIERRE	Professeur AgroParisTech	Dir. De thèse



Mes remerciements s'adressent en premier lieu à Lise Jouanin et Catherine Lapierre pour m'avoir accueillie dans leurs équipes pendant ces trois années. Je remercie également à Richard Sibout pour avoir participé à l'encadrement de cette thèse et grâce à qui *Brachypodium* est en phase de devenir une ressource essentielle de l'INRA de Versailles.

Je tiens à remercier tous les autres membres, anciens ou présents, de l'équipe paroi secondaire : Olivier Darracq, Nathalie Demont-Caulet, Nicolas Oria, Sébastien Antelme, Davy Baratiny, Julien Mazel, Serge Berthet, Sylvain Legay, Johanne Thévenin, Yin Wang, Aurélie Scagnelli et Marie Séverine Ho Yue Kuang et pour leur contribution à ce projet, qu'elle ait été d'ordre technique ou morale, ou même les deux ! Je souhaite remercier tout particulièrement Philippe Le Bris pour l'aide précieuse qu'il m'a apportée depuis son arrivée dans le laboratoire.

Merci également à tous les stagiaires qui sont passés par le laboratoire 255, Audrey Cou, Manon Jattiot, Alexia Guillebaux, Emilie Bernard, Valérie Cornuault, et tous les autres qui ne sont passés que brièvement mais ont su s'intégrer de façon toujours très sympathique.

Sans oublier les équipes et les collègues ayant participé activement à ce projet, notamment l'équipe Régénération et l'équipe Lignine et Tanin, et en particulier Oumaya Bouchabke-Coussa, Camille Soulhat, Laurent Cézard, Frédéric Légée et Brigitte Pollet. Un grand merci également à l'équipe d'Abdel Bendamane au génopôle d'Evry, grâce à qui la collection de mutants a pu voir le jour, et à Marion Dalmais pour la mise au point et l'exploitation du TILLING sur cette collection, merci à nos voisins de l'équipe développement et qualité des graines pour les analyses de phénols solubles. Merci Leonardo Gomez et Simon McQueen Mason d'avoir mis au point le système automatisé d'analyse de saccharification et d'y avoir analysé nos échantillons.

Ce qui nous amène au projet RENEWALL et à tous ses membres que je remercie pour l'accueil chaleureux et les discussions constructives lors des réunions auxquelles j'ai eu la chance de participer.

Merci à Patrick Grillot, Fabrice Petitpas, Philippe Maréchal, Bruno Letarnec et Hervé Ferry pour avoir pris soin de mes plantes en serre et à Halima Morin de la plateforme de cytologie pour ses conseils son aide.

Je souhaite remercier particulièrement les « voisins » du deuxième étage du bâtiment 2 ou un peu plus éloignés, qui ont bien souvent été de bon conseil et qui ont toujours constitué un soutien sans faille quelles que soient les situations, dans le désordre Susana, Michael, Volker, Alice, Jozef, Julien, Hélène, Adeline, Amel...

Enfin, merci à Claire Halpin, David Pot, et Michel Dron pour leur participation au jury, et surtout merci à Jacqueline Grima-Pettenati et Simon Hawkins d'avoir accepté la charge d'être rapporteurs de ce manuscrit.



Titre : Identification de gènes impliqués dans la voie de biosynthèse des monolignols chez *Brachypodium distachyon*.

Résumé :

La récente définition de *Brachypodium distachyon* comme modèle des graminées en fait un organisme de choix pour l'étude de leur paroi cellulaire, en particulier dans le cadre de leur utilisation comme matière première renouvelable pour le bioéthanol de seconde génération. Les lignines, dont les trois unités (H, G et S) proviennent de la polymérisation des monolignols, sont associées aux acides hydroxycinnamiques dans la paroi des céréales et représentent l'obstacle majeur à l'exploitation industrielle de la biomasse lignocellulosique. L'acquisition de connaissances sur les mécanismes dirigeant leur mise en place leur organisation permettrait d'identifier des facteurs modulant les rendements de production qui y sont associés. Quatre familles de gènes ont été étudiées et l'implication dans la voie de biosynthèse des monolignols de trois gènes a été montrée : *BdF5H2* possède une activité férulate-5-hydroxylase permettant la synthèse des *O*-Méthyltransférases et sa perte partielle de fonction cause une diminution de la quantité de lignine, la modification du rapport S/G et une baisse de quantité d'acide *p*-coumarique dans deux lignées mutantes indépendantes. Enfin, *BdCAD1* est l'isoforme principale des alcools cinnamylique déshydrogénases : sa perte de fonction dans deux lignées indépendantes cause la diminution de la quantité globale de lignine et d'acide *p*-coumarique, une baisse du rapport S/G ainsi que l'accumulation de sinapaldéhyde. Par ailleurs ces deux lignées présentent des rendements de saccharification augmentés de plus d'un quart par rapport au sauvage.

Title : Identification of genes involved in the biosynthesis of monolignols in *Brachypodium distachyon*

Abstract :

*Brachypodium distachyon* was recently adopted as an experimental model for grass species. As such, it is used to study grass cell wall, in particular in the context of their use as renewable feedstock for the production of second generation bioethanol. Lignins are polymers of three main units (H, G and S) originating from the polymerization of monolignols, and are linked to hydroxycinnamic acids in grasses. They constitute the main bottleneck to industrial processes targeting lignocellulosic biomass and improving the understanding of the mechanisms directing their structure and deposition could lead to the identification of the factors modulating associated production yields. Four gene families were studied and the involvement of three genes in the monolignols biosynthetic pathway was shown: *BdF5H2* displays a ferulate-5-hydroxylase activity enabling the synthesis of the S lignin units, *BdCOMT3* is the main caffeic acid *O*-methyltransferase and its partial loss of function in two independent mutant lines leads to the reduction of lignin content, the modification of the S/G units ratio and a decrease in *p*-coumaric acid accumulation. *BdCAD1* is the main cinnamyl alcohol dehydrogenase isoform: its loss of function in two independent mutant lines results in a decrease in lignin content and of the S/G ratio and the accumulation of sinapaldehyde. Moreover, these two lines display significantly increased saccharification yields.



## Sommaire général

<b>INTRODUCTION GENERALE .....</b>	<b>1</b>
<b>Chapitre I : Synthèse bibliographique .....</b>	<b>3</b>
<b>1. Les biocarburants.....</b>	<b>3</b>
1.1. Description .....	3
1.2. Le bioéthanol de seconde génération .....	5
1.3. Amélioration du rendement de production.....	6
1.4. Intérêt d'un modèle d'étude .....	8
<b>2. Brachypodium distachyon.....</b>	<b>9</b>
2.1. Description .....	9
2.2. Répartition géographique .....	11
2.3. Croissance et développement .....	12
2.4. Outils de travail disponibles.....	14
<b>3. Organisation de la paroi lignocellulosique végétale .....</b>	<b>16</b>
3.1. La lamelle moyenne .....	17
3.2. La paroi primaire .....	17
3.3. La paroi secondaire .....	19
<b>4. Constituants de la paroi végétale .....</b>	<b>20</b>
4.1. Protéines pariétales.....	20
4.1.1. Protéines structurales.....	20
4.1.2. Enzymes de modification pariétale .....	21
4.2. Silice.....	22
4.3. Pectines.....	23
4.4. Cellulose.....	24
4.5. Hemicelluloses .....	26
4.6. Acides hydroxycinnamiques .....	27
4.7. Lignines .....	29
4.7.1. Lignification .....	29
4.7.2. Description des lignines .....	29
4.7.3. Biosynthèse des monolignols.....	32
4.7.3.1. Etapes communes de la voie des phénylpropanoïdes.....	33
4.7.3.2. Voie spécifique aux monolignols .....	34
4.7.3.3. Transport et Polymérisation des monolignols .....	40
4.7.3.4. Régulation de la biosynthèse des monolignols.....	43
<b>Objectifs du travail de thèse.....</b>	<b>45</b>
<b>CHAPITRE II : Disrupting the cinnamyl alcohol dehydrogenase 1 (BdCAD1) gene leads to altered lignification and improved saccharification in Brachypodium distachyon.....</b>	<b>47</b>
<b>Annexe au chapitre II : .....</b>	<b>68</b>
<b>1. Expression en fonction du cycle circadien .....</b>	<b>68</b>
<b>2. Profils d'expression en fonction des tissus .....</b>	<b>69</b>
<b>3. Profil d'expression par hybridation <i>in situ</i> .....</b>	<b>70</b>





### **Chapitre III : Identification de l'acide cafféique *O*-méthyltransférase impliquée dans la voie de biosynthèse des monolignols..... 73**

1.	Détermination du gène candidat par alignements de séquences.....	75
2.	Etude phylogénétique.....	77
3.	Expression en fonction du cycle circadien .....	77
4.	Profils d'expression en fonction des tissus .....	78
4.1.	Bradinet .....	78
4.2.	Etudes d'expression par RT-PCR .....	78
5.	Complémentation fonctionnelle .....	79
5.1.	Analyses des lignines .....	79
5.2.	Composés phénoliques solubles.....	81
6.	Identification de mutants par TILLING.....	82
6.1.	Recherche .....	82
6.2.	Prédictions des effets des mutations.....	84
6.2.1.	Prédiction par l'outil SIFT .....	84
6.2.2.	Prédiction par modélisation.....	85
7.	Etude de la lignification des mutants.....	88
8.	Conclusion et perspectives.....	90

### **Chapitre IV : Identification de la cafféoyl-coenzyme A 3-*O*-méthyltransférase impliquée dans la voie de biosynthèse des monolignols. 92**

1.	Détermination du gène candidat par alignements de séquences.....	94
2.	Etude phylogénétique.....	95
3.	Expression en fonction du cycle circadien .....	96
4.	Profils d'expression en fonction des tissus .....	96
4.1.	RT-PCR.....	96
4.2.	Profils Bradinet .....	97
5.	Complémentation fonctionnelle .....	97
6.	Identification de mutants <i>Bdccoamt1</i> par TILLING .....	98
6.1.	Prédiction de l'effet des mutations.....	99
6.2.	Analyse des lignines des lignées identifiées .....	100
7.	Conclusion et perspectives.....	102

### **Chapitre V : Identification de la férulate-5-hydroxylase impliquée dans la voie de biosynthèse des monolignols..... 104**

1.	Détermination du gène candidat par alignements de séquences.....	106
2.	Etude phylogénétique.....	107
3.	Expression en fonction du cycle circadien .....	108
4.	Profils d'expression .....	108
4.1.	RT-PCR.....	108



4.2.	Bradinet .....	109
<b>5.</b>	<b>Complémentation fonctionnelle .....</b>	<b>109</b>
5.1.	Analyses des lignines des tiges matures .....	109
5.2.	Composés phénoliques solubles .....	111
<b>6.</b>	<b>Conclusion et perspectives .....</b>	<b>111</b>
<b>Discussion générale et perspectives .....</b>		<b>113</b>
<b>Références bibliographiques .....</b>		<b>118</b>
<b>ANNEXES .....</b>		<b>130</b>
<b>MATERIEL ET METHODES.....</b>		<b>131</b>
<b>1.</b>	<b>Protocoles de bioinformatique .....</b>	<b>131</b>
1.1.	Recherche de séquences .....	131
1.2.	Formatage et traduction.....	132
1.3.	Alignements de séquences .....	132
1.4.	Arbres phylogénétiques.....	133
1.5.	Prédiction de structures protéiques 3D.....	133
1.6.	Visualisation et modification des protéines modélisées.....	134
1.7.	Prédiction de l'impact des mutations sur l'activité d'une protéine d'intérêt.....	135
<b>2.</b>	<b>Protocoles de biologie moléculaire.....</b>	<b>135</b>
2.1.	Extraction d'ADNg au CTAB (Arabidopsis et Brachypodium, tous organes) 135	
2.2.	Extraction d'ARN (Arabidopsis et Brachypodium).....	136
2.3.	Rétrotranscription.....	138
2.4.	PCR .....	138
2.5.	Clonage GATEWAY® (Invitrogen) .....	139
2.6.	Extraction plasmidique par lyse alcaline.....	142
2.7.	Transformation d'Arabidopsis thaliana.....	143
2.8.	Stérilisation de graines d' <i>Arabidopsis</i> .....	144
2.9.	Sélections des transformants primaires ( <i>Arabidopsis</i> T1).....	145
2.10.	Sélection des transformants secondaires ( <i>Arabidopsis</i> T2).....	145
2.11.	Sélection de transformants tertiaires ( <i>Arabidopsis</i> T3).....	146
<b>3.</b>	<b>Protocoles de cytologie .....</b>	<b>147</b>
3.1.	Coupes transversales au vibratome .....	147
3.2.	Coloration de lignines au phloroglucinol (réactif de Wiesner) .....	147
3.3.	Coloration de Maüle.....	148
3.4.	Hybridations <i>in situ</i> .....	149
<b>4.</b>	<b>Protocoles d'analyses de lignines .....</b>	<b>151</b>
4.1.	Protocole de thioacidolyse .....	151
<b>5.</b>	<b>Protocoles de biochimie .....</b>	<b>151</b>
5.1.	Extraction et dosage des protéines totales.....	151
5.2.	Test d'activité enzymatique CAD .....	152
<b>PREDICTIONS SIFT .....</b>		<b>153</b>



# INTRODUCTION GENERALE

L'augmentation des besoins en énergie, liée au développement économique et démographique mondial faisant face à la déplétion des carburants fossiles, a porté l'attention sur la production de biomasse dans le but de générer de l'énergie sous forme de biocarburants. En effet, les biocarburants de seconde génération procurés par les parois lignocellulosiques végétales sont envisagés comme une alternative efficace pour répondre à ces besoins, car présentant un impact limité sur les ressources alimentaires. Dans ce contexte, la culture de céréales est privilégiée, notamment au vu de la quantité de biomasse cultivable avec des apports limités, et *Brachypodium distachyon* ayant été adopté comme modèle d'étude des Poacées, c'est sur cet organisme que ce sont concentrés nos travaux de recherche. L'une des étapes les plus contraignantes du procédé de fabrication de biocarburants est la saccharification, ou conversion des sucres complexes de la paroi en sucres simples fermentescibles. C'est sur ce point que se concentre le projet européen RENEWALL, projet finançant ces travaux de thèse, dont le but principal est l'identification, à plusieurs niveaux, des facteurs permettant l'amélioration des rendements de saccharification. L'étape de saccharification est en grande partie conditionnée par deux facteurs : la quantité de sucres disponibles dans la biomasse, et la facilité de l'accès à ces sucres par les enzymes de saccharification. Ce dernier facteur est fortement influencé par la quantité et la structure des lignines présentes dans la paroi végétale. Les lignines proviennent de la polymérisation de trois principaux monolignols, produits à partir de la phénylalanine par une voie de biosynthèse complexe dont les modifications ont des effets parfois radicaux sur la composition pariétale.



S'inscrivant dans le projet RENEWALL et dans la thématique de l'équipe « paroi secondaire et lignine » de l'IJPB à l'INRA de Versailles, ce travail de thèse a consisté en l'identification de gènes impliqués dans la voie de biosynthèse des monolignols chez *Brachypodium distachyon* et l'analyse des effets de modifications de l'expression de ces gènes sur la structure des lignines et les rendements de saccharification des parois des lignées étudiées. Les cinq familles de gènes étudiées sont celles impliquées dans les dernières étapes de la voie de biosynthèse, afin de limiter autant que possible à la lignification les effets de leurs modifications.

Le chapitre I consiste en une revue bibliographique présentant un état des lieux des connaissances sur l'espèce étudiée et les raisons du choix de *Brachypodium* comme espèce modèle, sur la constitution de la paroi et la lignification et sur les raisons du développement de l'industrie des biocarburants.

Les chapitres suivants aborderont les travaux d'identification, au sein de chaque famille étudiée, des gènes impliqués de façon majeure dans la voie de biosynthèse.

Le chapitre II porte sur l'identification du gène *CAD* et l'analyse de mutants et sera présenté sous forme d'un article scientifique.

Le chapitre III porte sur l'identification du gène *COMT* et l'analyse de mutants.

Le chapitre IV porte sur l'identification du gène *F5H*.

Le chapitre V porte sur l'identification du gène *CCoAOMT*.

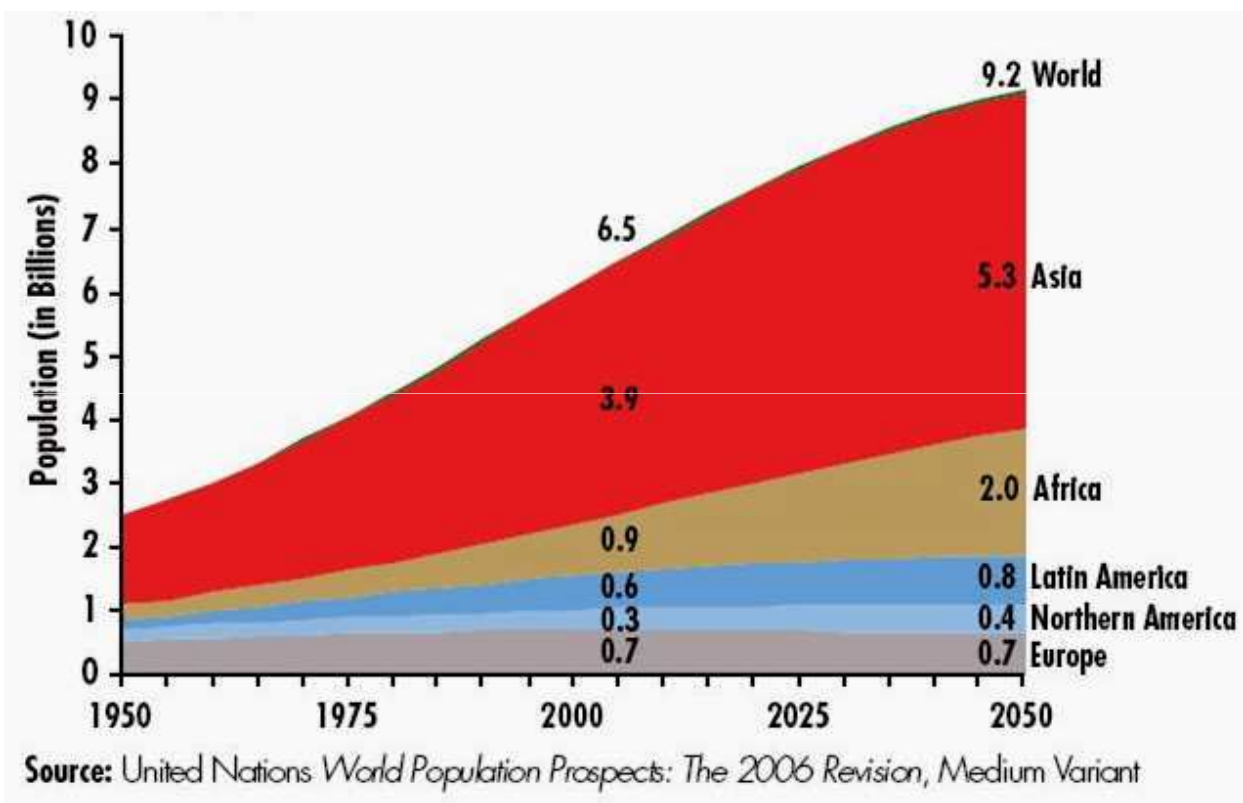
Le chapitre VI présentera les conclusions et les perspectives de ces travaux.





Chapitre I :

SYNTHESE BIBLIOGRAPHIQUE



**Figure 1.** Croissance et pronostic d'évolution de la population mondiale entre 1950 et 2050. Les populations des différents continents sont chacune représentées par une couleur.

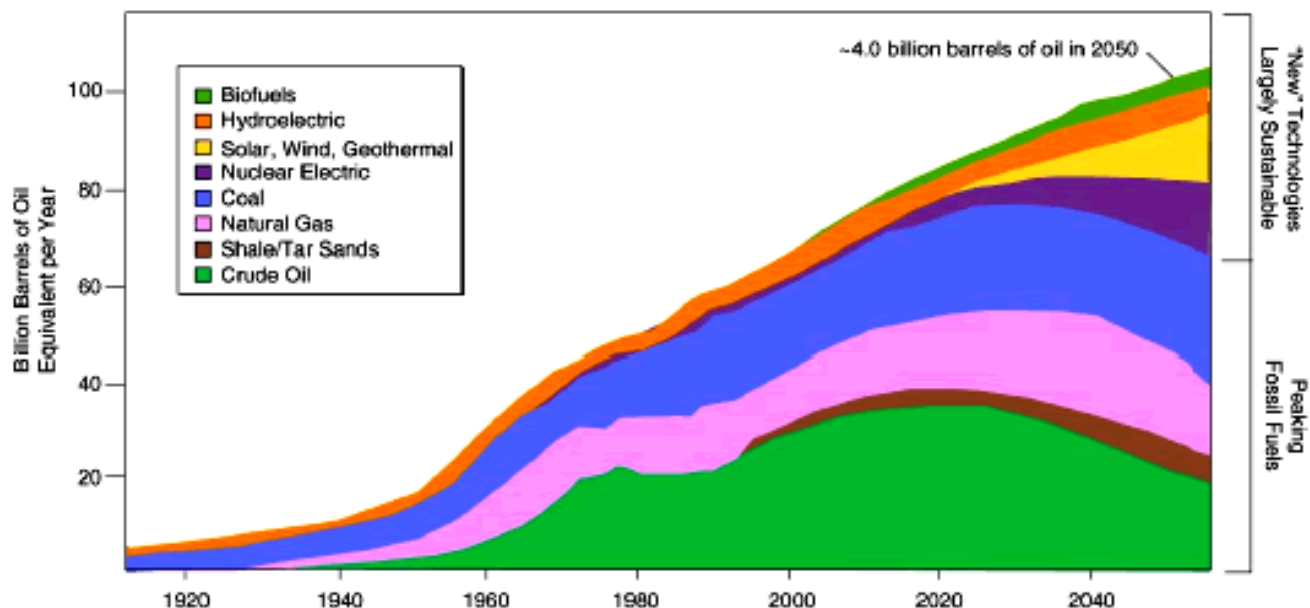
# Chapitre I : Synthèse bibliographique

## 1. Les biocarburants

### 1.1. Description

Les besoins énergétiques à l'échelle mondiale n'ont cessé d'augmenter ces dernières décennies et sont continuellement en expansion. Les trois principaux facteurs de cette augmentation sont la croissance démographique mondiale, la croissance économique (bien que celle-ci soit quelque peu ralentie depuis 2008...) et l'émergence des pays en développement. En effet, la population mondiale qui était de 6,5 milliards en 2005 est prédite de dépasser les 9 milliards en 2050 (Fig. 1). Cette croissance sera la plus importante en Afrique et en Asie, notamment en Chine et en Inde, où elle est associée à une augmentation du niveau de vie. Ceci implique des demandes de plus en plus importantes en énergies nécessaires au transport, à l'agriculture et aux industries, ainsi qu'au chauffage et à l'éclairage des habitations privées et des bureaux. A ce jour, la grande majorité des ressources énergétiques provient de ressources fossiles que sont le charbon, le pétrole et le gaz naturel. Cependant, ces ressources ne sont pas renouvelables et risquent de s'épuiser si leur consommation continue au rythme actuel. D'autres sources d'énergies sont actuellement exploitées, les principales étant l'énergie nucléaire, également non renouvelable, et l'énergie hydroélectrique. Les énergies alternatives, utilisées dans des proportions bien plus faibles, comprennent l'énergie solaire convertie par des cellules photovoltaïques, les énergies provenant du vent, des marées ou des vagues converties en électricité par des turbines, l'énergie géothermique et enfin les bioénergies provenant de matières premières végétales brûlées ou transformées pour convenir à d'autres utilisations. Parmi ces énergies

## World Energy Demand—Long-Term Energy Sources



Sources: Lynn Orr, *Changing the World's Energy Systems*, Stanford University Global Climate & Energy Project (after John Edwards, American Association of Petroleum Geologists); SRI Consulting.

**Figure 2.** Consommation et pronostic d'évolution des différentes sources d'énergies entre 1910 et 2060, estimées en équivalent de milliard de baril de pétrole. Chaque source d'énergie est représentée par une couleur.

**Tableau 1.** Comparaisons énergétique et écologique des différents types de biocarburants et leurs sources de biomasse. (Plants to power : bioenergy to fuel the future Yuan 2008 Cell press)

Platforms <sup>a</sup>	Feedstock <sup>a</sup>	NEB <sup>b</sup> GJ/ha/yr	NER <sup>b</sup>	CO <sub>2</sub> balance	Annual feedstock	Estab-lishment	Germ-plasm	Agricul. practice <sup>d</sup>	Ecological benefits
Ethanol from starch or sucrose	Maize	10–80	1.5–3.0	Positive	Yes	+++ <sup>c</sup>	+++	+++	+
	Sugarcane	55–80	3.0–4.0	Positive	No	+++	+++	+++	+
	Sugar beet	40–100	2.5–3.5	Positive	Yes	+++	++	+++	+
	Sweet sorghum	85–300	5–10	Positive	Yes	+++	++	++	++
Ethanol from lignocellulosic feedstocks	Miscanthus	250–550	15–70	Possibly negative	Yes/No	+	+	+	+++
	Switchgrass	150–500	10–50	Possibly negative	No	+	+	+	+++
	Poplar	150–250	10–20	Possibly negative	No	+	++	++	+++
Biodiesel	Soybean	–20–10	0.2–0.6	Positive	Yes	++	+++	+++	+
	Canola	–5–2	0.7–1.0	Positive	Yes	+++	+++	+++	+
	Sunflower	–10–0	0.3–0.9	Positive	Yes	+++	++	+++	+

<sup>a</sup> Multiple platforms and crops are compared in a synthesis integrating information from multiple studies.

<sup>b</sup> Abbreviations: NEB, energy balance; NER, net energy ratio, which is the ratio of output to input energy needed to produce a fuel from a feedstock.

<sup>c</sup> Favorable features are indicated by + symbols, with +++ being the most favorable.

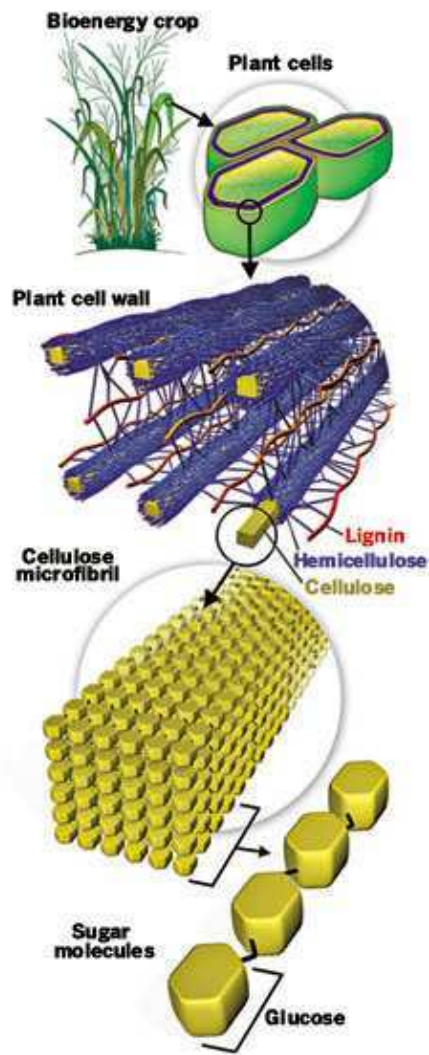
<sup>d</sup> Agricul. Practice, agricultural practice: how advanced is the current status of farming, harvesting, and processing.

renouvelables, les seules à pouvoir efficacement remplacer les carburants fossiles employés pour le transport sont les bioénergies sous forme de biodiesel ou de bioéthanol (Fig. 2).

La rentabilité économique et environnementale d'un biocarburant peut être mesurée par son bilan énergétique (NEB, net energetic balance) exprimé en GJ/ha/an. Le bilan carbone permet d'évaluer l'impact écologique de ces carburants, en rapportant à un équivalent carbone tous les processus physiques liés à leur production, notamment la production de gaz à effet de serre. Un bilan carbone neutre correspond à la capacité de recyclage du total des émissions de gaz à effet de serre. Un bilan carbone positif sera considéré comme produisant plus de ces polluants qu'il n'en est recyclable. Idéalement, les énergies renouvelables sont supposées présenter un bilan carbone négatif.

Le biodiesel, employé en mélange avec le diesel, est fabriqué à partir des lipides extraits des graines de plantes oléagineuses tels que le soja (*Glycine max*), le tournesol (*Helianthus annuus*) ou le colza (*Brassica napus*) puis transestérifiés par le méthanol. Malgré son efficacité et utilisation déjà répandues, ce biocarburant possède les inconvénients majeurs d'entrer en compétition directe avec les usages alimentaires des cultures dont ils proviennent et de présenter un mauvais bilan énergétique et un bilan carbone positif (Tableau 1).

Le bioéthanol est produit à partir de la fermentation des sucres fermentescibles extraits de la biomasse végétale. Dans le cas du bioéthanol de première génération, il s'agit des hexoses simples, majoritairement du glucose, contenus dans les tissus de réserve. Les sources principales de ces glucides sont l'amidon (polymère de glucoses liés en  $\alpha$ 1-4 et  $\alpha$ 1-6 facilement hydrolysable) issu des grains de maïs (*Zea mays*) ou d'autres céréales, et le saccharose (glucose et de fructose liés en  $\alpha$ 1-2) provenant de la betterave (*Beta vulgaris*) ou de la canne à sucre (*Saccharum officinarum*). Ces modes de production sont également en concurrence directe avec la production alimentaire et bien que leur bilan énergétique soit supérieur à celui du biodiesel, leur bilan carbone est toujours positif (Yuan, Tiller et al. 2008).



**Figure 3.** Localisation du glucose dans la paroi des plantes dédiées à la production de biocarburants (source: Genome Management Information System/ORNL à <http://pubs.acs.org/cen/coverstory/86/8633cover3.html> )

Les caractéristiques de ces deux biocarburants en font des solutions acceptables pour compenser l'épuisement des ressources fossiles, mais elles ne sont pas envisageables à long terme en l'état actuel de leur production. C'est pour ces raisons que les biocarburants de seconde génération sont actuellement considérés comme la solution alternative à long terme pour le remplacement des carburants fossiles (Tableau 1).

## 1.2. Le bioéthanol de seconde génération

Les biocarburants de seconde génération sont produits à partir de la biomasse lignocellulosique végétale (Fig. 3). La principale source de sucres fermentescibles est la cellulose, les hémicelluloses peuvent également être utilisées. Cependant la complexité de l'organisation des polymères pariétaux rend difficile l'accès à ces composants, surtout à cause de la présence de lignine qui perturbe l'action des enzymes de dégradation de la cellulose. En effet, le processus de production commence par un prétraitement thermique et chimique du matériel végétal, permettant de libérer les hémicelluloses et de faciliter l'accès à la cellulose. L'étape suivante est la saccharification des polysaccharides pariétaux, qui consiste en la conversion enzymatique des sucres complexes en sucres simples fermentescibles par des cellulases. Enfin, ces sucres sont fermentés en éthanol par l'action de bactéries ou de levures. Le choix des plantes productrices de la matière première lignocellulosique est à la base de l'efficacité du procédé. Des espèces de grande taille, comme Miscanthus ou l'herbe à éléphant (*Pennisetum purpureum*) permettent d'obtenir une grande quantité de biomasse. Mais ce n'est pas le seul critère de sélection et d'autres peuvent être pris en compte. Notamment, les plantes possédant un métabolisme en C4 auront tendance à utiliser plus efficacement l'eau et l'azote à leur disposition. De plus, les espèces pérennes remobilisent les nutriments de leurs parties aériennes vers leurs racines à la fin de leur saison de croissance. Ainsi, leur récolte épuise moins rapidement les ressources du sol. Un autre critère de sélection est leur adaptation aux





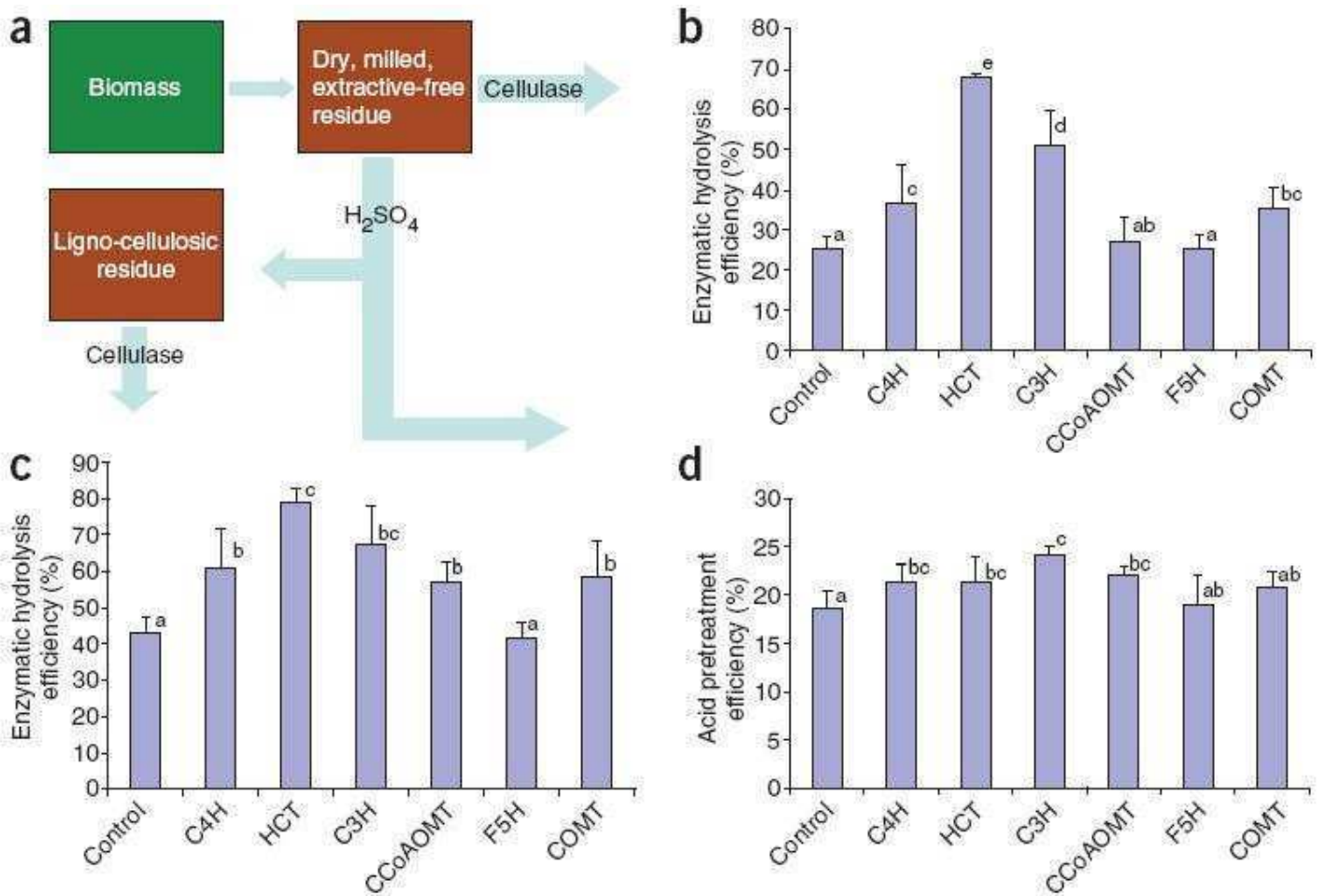
territoires proches des lieux de production et de consommation des biocarburants, leur transport affectant leur rentabilité énergétique et le bilan carbone de leur fabrication. Ainsi, un certain nombre de Poacées telles que le maïs, miscanthus ou encore le sorgho ont été suggérés comme plantes productrices de biomasse lignocellulosique dédiées, au moins en partie, à l'obtention de bioéthanol de seconde génération dans les régions tempérées (Yuan, Tiller et al. 2008).

Il existe cependant à l'heure actuelle quelques obstacles majeurs à la fabrication de ces biocarburants :

- la mise en place des infrastructures de production est en cours mais leur optimisation et la construction d'un grand nombre de raffineries reste nécessaire pour espérer pouvoir répondre, à terme, à la demande envisagée.
- L'utilisation des infrastructures déjà existantes pour le transport de carburants fossiles ne pourra être utilisée que pour du bioéthanol fortement dilué car elles n'ont pas été prévues pour supporter son acidité et risqueraient d'être très rapidement endommagées. Pour le transport sous forme non diluée, les canalisations des pipelines ou les citernes devront être adaptées (USDE Biofuels Joint Roadmap, 2006).
- Le rendement de production actuel n'est pas encore à son niveau optimum, grandement à cause de la récalcitrance à la saccharification de la biomasse lignocellulosique.

### 1.3. Amélioration du rendement de production

Bien que la digestibilité ait été un point d'étude des fourrages, les processus de sélection en vue d'applications agronomiques n'ont que très peu visé l'optimisation des graminées dans le but d'exploiter leur biomasse lignocellulosique. Ces végétaux possédant des parois complexes, de nombreux travaux sont en cours pour identifier les facteurs pouvant être visés



**Figure 4.** Rendements de saccharification de lignées RNAi présentant une baisse d'activité des enzymes de la voie de biosynthèse des monoligols par rapport au sauvage chez la luzerne. (Chen et Dixon 2007)

pour améliorer les rendements de production de bioéthanol. A titre d'exemple, à l'heure actuelle la plante au rendement le plus élevé est Miscanthus avec entre 4600 et 12400 Litres de bioéthanol par hectare mais les tiges de maïs ne produisent en moyenne que 900L/ha (Somerville, Youngs et al. 2010).

Le projet européen RENEWALL recense les points clés pour ces recherches qui sont :

- La compréhension de la biosynthèse et la chimie des polymères pariétaux, ainsi que leurs interactions résultant en leur récalcitrance à l'hydrolyse
- L'identification de gènes impliqués dans la biosynthèse des parois, pouvant être manipulés pour améliorer le rendement de saccharification
- L'expression dans les plantes d'enzymes modifiant ou dégradant les parois et les rendant plus susceptibles à la saccharification
- L'amélioration des prétraitements pour augmenter la transformation des matières premières
- La mise en place de stratégies de cultures des plantes possédant un rendement de saccharification augmenté.

La lignine est l'un des facteurs impactant négativement la conversion des parois lignocellulosique (Lynd, Cushman et al. 1991) et il a notamment été montré que la quantité et la composition en lignines dans les parois est fortement corrélée à leur facilité à être dégradées (Chapple, Ladisch et al. 2007). La biosynthèse et la polymérisation des monolignols sont donc des cibles privilégiées pour ces manipulations. Il existe de nombreux mutants dans cette voie chez plusieurs espèces dont les modifications des parois permettent une amélioration de la saccharification. Ceci est très bien illustré par l'étude de Chen et Dixon sur la luzerne (Chen and Dixon 2007) (Fig. 4).

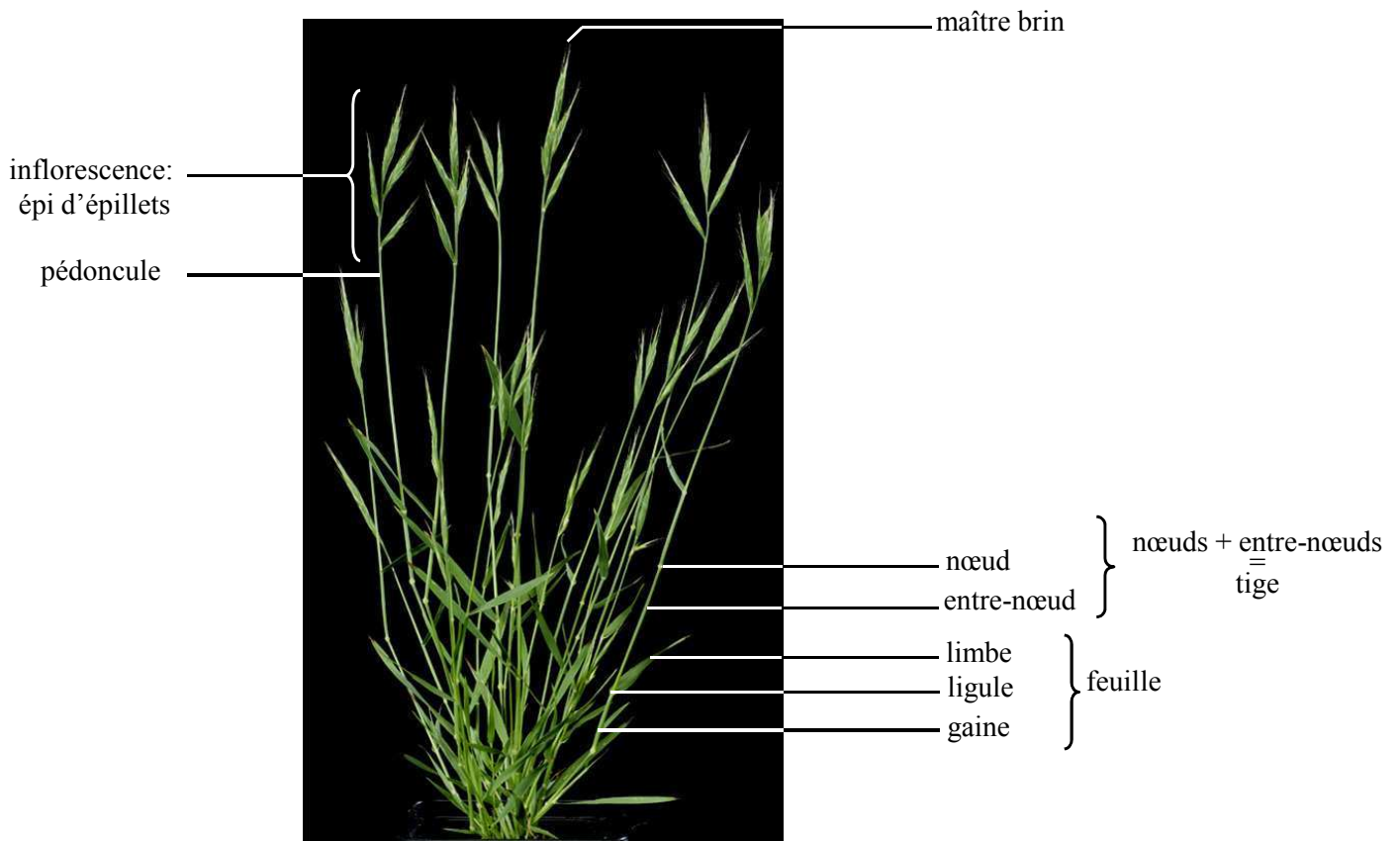
Un autre avantage de l'utilisation de Poacées pour la production de bioéthanol de seconde génération vient des structures de leurs polymères de lignines. En effet, la majeure partie des



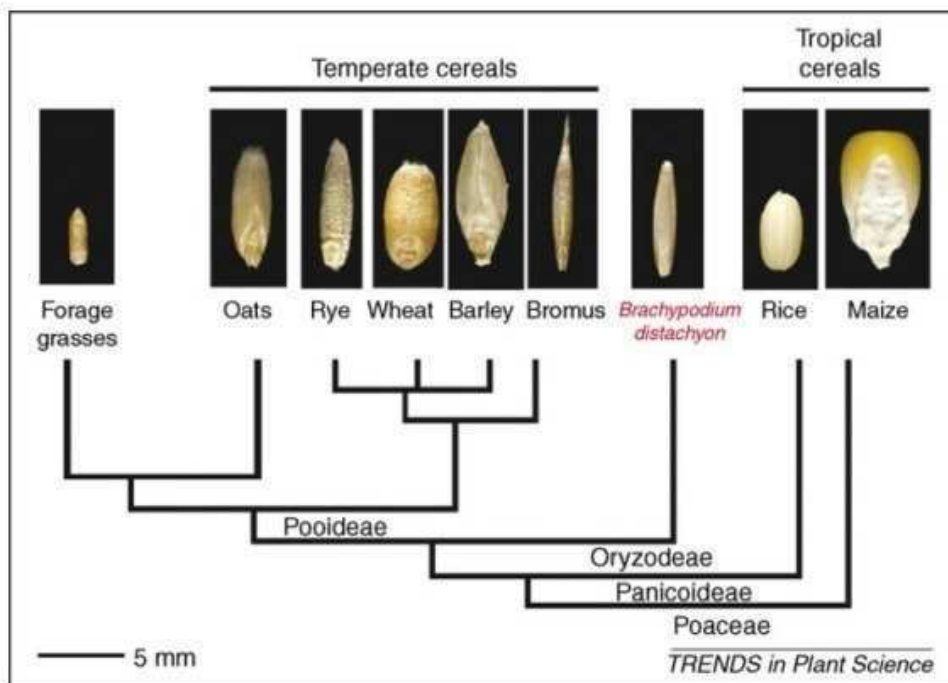
monomères y sont liés par des liaisons résistantes et l'ensemble de la structure contient plus de ramifications. Ceci a un impact direct sur l'architecture de la paroi, où les lignines sont moins adsorbées (c'est-à-dire fixées par liaisons faibles du type Wan Der Waals) aux polysaccharides que dans le cas d'un polymère plus linéaire. Cette organisation a pour conséquence de faciliter l'accès aux enzymes de dégradation pour la saccharification.

#### 1.4. Intérêt d'un modèle d'étude

Bien que certaines dicotylédones comme le peuplier (genre *Populus*) soient candidates pour la production de biomasse lignocellulosique dédiée à la production de bioéthanol de seconde génération, les espèces envisagées sont en grande majorité des Poacées (Rapport Doe, 2006). La composition des parois lignocellulosiques variant de façon importante entre les Poacées et les dicotylédones, cela souligne la nécessité de travailler sur un modèle dont la structure pariétale est similaire à celle des espèces sur lesquelles les applications à l'échelle agronomique sont prévues (Gomez 2008, Opanowicz 2008). L'état des connaissances sur la paroi des céréales est déjà avancé, cependant l'identification d'ensembles de caractères spécifiques permettant la modulation de l'accès aux polysaccharides pariétaux conjointement à l'augmentation des proportions de ces composés saccharifiables reste à affiner. Ainsi la nécessité d'un modèle d'étude simple et disposant de toutes les ressources permettant les études génomiques et fonctionnelles à grande échelle s'est imposé. *Brachypodium* remplissant les critères d'une espèce modèle, il a été désigné comme modèle et est à présent couramment utilisé pour les études sur les parois lignocellulosiques.



**Figure 5.** Anatomie de *Brachypodium distachyon*. En moyenne, les plantes peuvent atteindre 45 cm dans de bonnes conditions de culture.



**Figure 6.** Relations phylogénétiques entre *Brachypodium* et des céréales d'importance économique (Opanowicz *et al.* 2008)

## 2. *Brachypodium distachyon*

*Brachypodium* a été envisagé comme modèle d'étude depuis une décennie (Draper 2001) et adopté depuis quelques années en raison de ses nombreuses caractéristiques permettant un travail à grande échelle sur plusieurs générations chaque année.

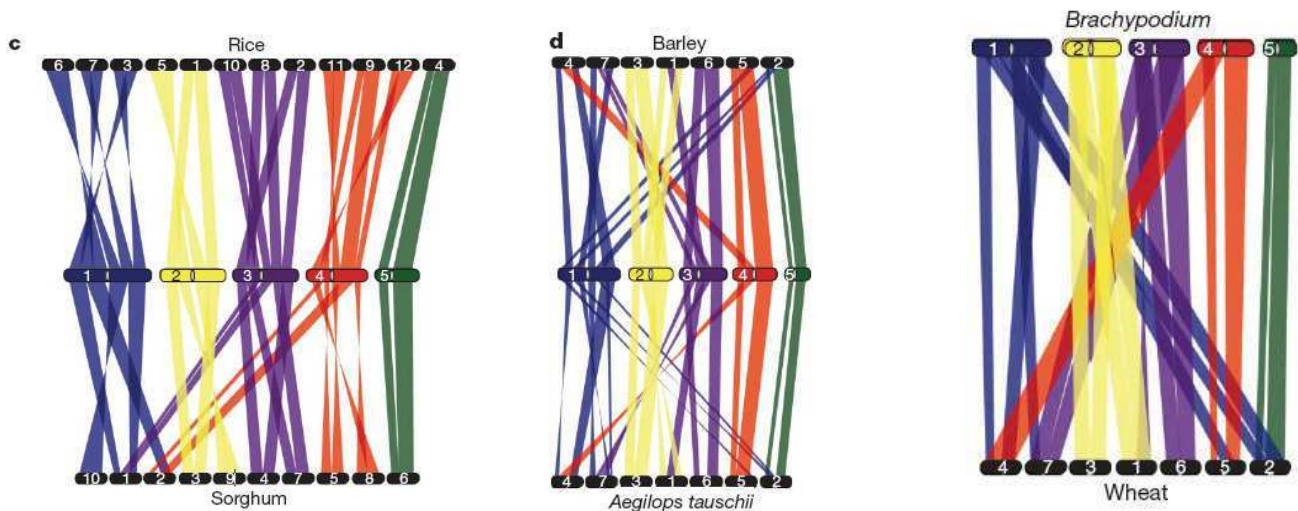
### 2.1. Description

*Brachypodium distachyon* (L.) Beauv. (Fig. 5), ou *Brachypodium*, communément appelé Brachypode à deux épis (ou en anglais purple false brome), est une angiosperme monocotylédone annuelle de l'ordre des *Poales*, de la famille des *Poaceae* (appelées autrefois graminées), appartenant à la sous-famille des *Pooideae* et à la tribu ou infra-famille des *Brachypodieae* issue de la divergence de celle des *Poeae* et de celle des *Triticeae* il y a 35 à 40 millions d'années (Catalán, Kellogg et al. 1997). La famille des Poacées compte des céréales de grande culture telles que le blé (*Triticum aestivum*), le riz (*Oryza sativa*), le maïs (*Zea mays*), l'orge (*Hordeum vulgare*), l'avoine (*Avena sativa*) ou le seigle (*Secale cereale*), ainsi que des céréales fourragères telles que le ray-grass (*Lolium multiflorum* ou *perenne*), les bromes (*Bromus cartharticus* et *sitchensis*) ou la fétuque (*Festuca arundinacea*) et des céréales destinées à la production de biomasse telles que miscanthus (*Miscanthus giganteus*) ou le panic (*Panicum virgatum*). De nombreuses analyses phylogénétiques, notamment sur les BAC (Huo, Jiang et al. 2007) ou les séquences chloroplastiques (Catalán and Olmstead 2000) ont permis de situer *Brachypodium* entre le riz et les céréales de cultures tempérées (Fig. 6). Il est estimé que *Brachypodium* a divergé de son ancêtre commun avec le blé, il y a 32 à 39 millions d'années (Ma), de celui avec le riz il y a 40 à 53 Ma, et de celui avec le Sorgho il y a 45 à 60 Ma. D'après ces études, *Brachypodium* se situe donc entre les céréales de zones tempérées (telles que le blé) et celles de zones plus tropicales (comme le riz). Par ailleurs, il





s'agit d'une céréale possédant un métabolisme en C3, comme le blé, l'orge, l'avoine ou le riz, et non en C4, comme le maïs, la canne à sucre ou le sorgho. Cependant, les différences métaboliques ayant peu d'influence sur les composés structuraux, *Brachypodium* est un organisme modèle valable également pour la recherche sur les plantes en C4 dans certains domaines, notamment ceux ayant pour application la génération de biomasse constituée de paroi lignocellulosique dédiée à la production de biocarburants. *Brachypodium* est aussi un modèle utilisé pour les études sur la floraison et les graines en raison de sa similarité avec les espèces cultivées pour leur grain comme le riz et le blé (Laudencia-Chingcuanco and Vensel 2008; Guillon, Bouchet et al. 2011; Quraishi, Murat et al. 2011). Par rapport aux céréales d'intérêt agronomique telles que le maïs ou le blé, *Brachypodium* présente de nombreux avantages. Du point de vue des besoins en infrastructures, les plantes sont de plus petite taille, atteignant en moyenne 20 cm et au maximum 50 cm suivant les conditions de culture, *Brachypodium* peut être cultivé avec une densité atteignant 1000 plantes au mètre carré. En comparaison, les cultures de la plante modèle *Arabidopsis* peuvent atteindre 2000 plantes par m<sup>2</sup>. Cependant, les cultures de blé atteignent une densité maximale de 50 plantes par m<sup>2</sup>, ce qui est encore nettement supérieur aux densités de cultures du riz (36 plantes par m<sup>2</sup>), du sorgho (13 plantes par m<sup>2</sup>), du maïs ou du panic (*Panicum virgatum*) (6 plantes par m<sup>2</sup>), rendant contraignantes en terme d'espace les études sur un grand nombre de plantes ou lignées. Par ailleurs, le temps de génération de *Brachypodium* peut être compris entre 8 et 12 semaines dans les conditions de cultures adéquates et permet de travailler sur jusqu'à 4 générations par an comme pour le modèle *Arabidopsis*. Plusieurs générations peuvent être obtenues en une année pour le blé, le sorgho et le maïs, mais le riz et le panic ne peuvent généralement fournir qu'une seule génération par an. *Brachypodium* est une espèce relativement peu exigeante en matière d'intrants, cependant ceci ne constituerait un avantage uniquement en comparaison avec le riz qui est une céréale nécessitant des conditions de



**Figure 7.** Relations synténiques entre *Brachypodium* et d'autres céréales (The *Brachypodium* Genome Initiative, 2010): relations entre les chromosomes de *Brachypodium* et d'autres espèces de céréales. Les ensembles de relations orthologues colinéaires sont représentées par des bandes de couleur suivant chaque chromosome de *Brachypodium* (bleu= Ch1; jaune= Ch2; violet= Ch3; rouge= Ch4; vert= Ch5) et les centromères sont représentés en blanc.

**Tableau 2.** Comparaison de caractéristiques entre *Brachypodium* et d'autres céréales et le modèle des dicotylédones *Arabidopsis thaliana* (Bevan et al., 2010).

Comparison of model and crop plants							
	<i>Brachypodium</i>	<i>Arabidopsis</i>	Rice	Sorghum	Wheat	Maize	Switchgrass
Height (cm)	15–20	15–20	100	170–320	50	155–215	200
Generation time (weeks)	8–12	8–12	30	17	12	14–20	26
Density (plants/m <sup>2</sup> )	1000	2000	36	13	50	6	6
Growth requirements	Simple	Simple	Demanding	Simple	Simple	Simple	Simple
Reproduction	Selfing	Selfing	Selfing	Outcrossing: self-compatible	Selfing	Outcrossing: self-compatible	Outcrossing: self-incompatible
Genome size (Mbp)	272 <sup>a</sup>	119 <sup>a</sup>	382 <sup>a</sup>	758 <sup>a</sup>	17 000	2048 <sup>a</sup>	2400–3200
Ploidy	2X	2X	2X	2X	6X	2X	4X–8X
Cell wall type <sup>b</sup>	Type II	Type I	Type II	Type II	Type II	Type II	Type II

<sup>a</sup> Assembled genome size.  
<sup>b</sup> Grass cell walls differ from dicot cell walls in the type of hemicellulose and the amounts of pectin, protein and phenolic compounds [32].

culture complexes (cette caractéristique du riz est la raison majeure de son impossibilité à devenir une espèce modèle). La reproduction par autofécondation de *Brachypodium* permet d'entretenir facilement des lignées homozygotes, et des techniques de croisements ont été mises au point pour les cas où ils sont nécessaires. Parmi les nombreuses accessions naturelles, Bd21-3 et Bd21, les plus couramment utilisées pour la recherche, sont diploïdes ce qui limite les redondances lors d'études génomiques. De plus, *Brachypodium* possède l'un des plus petits génomes parmi les céréales : ses 5 chromosomes ( $2n=10$ ) comportent 25532 gènes répartis sur environ 272 Mb (Vogel, Garvin et al. 2010), constituant un génome compact possédant un fort taux de recombinaison (Huo, Vogel et al. 2009) favorisant le développement d'outils de travail tels que le séquençage ou le clonage positionnel. Les accessions polyploïdes pourraient également trouver leur intérêt dans l'élargissement à l'étude de céréales polyploïdes dont l'exemple agronomique majeur est le blé. Le degré de colinéarité du génome de *Brachypodium* avec ceux des céréales cultivées (Fig. 7) a permis de renforcer l'utilisation de cet organisme modèle (Vogel, Garvin et al. 2010).

Si la plante modèle *Arabidopsis* présente les mêmes avantages, elle a l'inconvénient de posséder des parois de type I, contrairement à *Brachypodium* ou aux céréales de grandes culture ayant en commun des parois de type II (aspects développés en paragraphe 2.2). En résumé, le faible espace nécessaire, la courte durée de chaque génération et la flexibilité des modes de reproduction ont été des éléments clés dans le choix de *Brachypodium* en tant qu'espèce modèle (Tableau 2).

## 2.2. Répartition géographique

*Brachypodium* est originaire du bassin méditerranéen (Draper 2001; Hasterok, Draper et al. 2004; Vogel, Gu et al. 2006). En particulier, l'accession naturelle Bd21 provient d'Irak. Cependant certaines accessions sont présentes jusqu'en Inde, en Afrique de l'Est, sur le

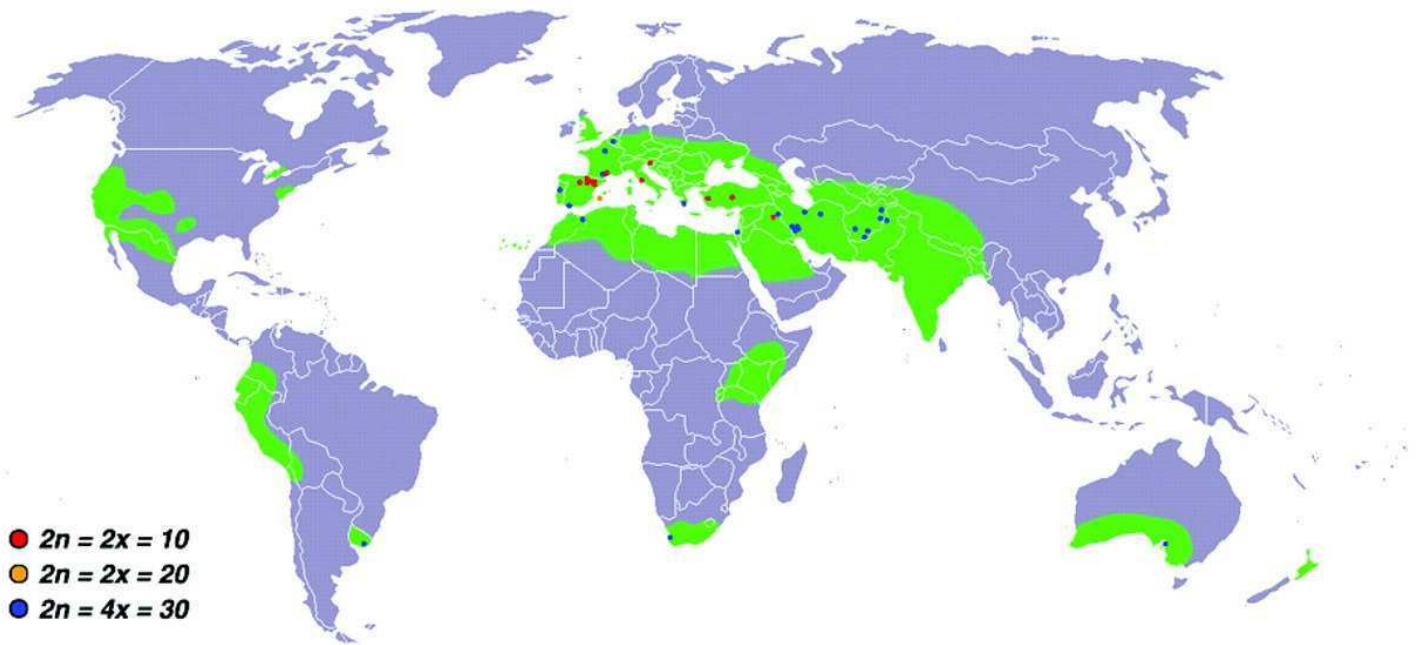


Figure 8. Répartition de *Brachypodium distachyon* dans le monde et ploïdie des accessions présentes. (crops.org)

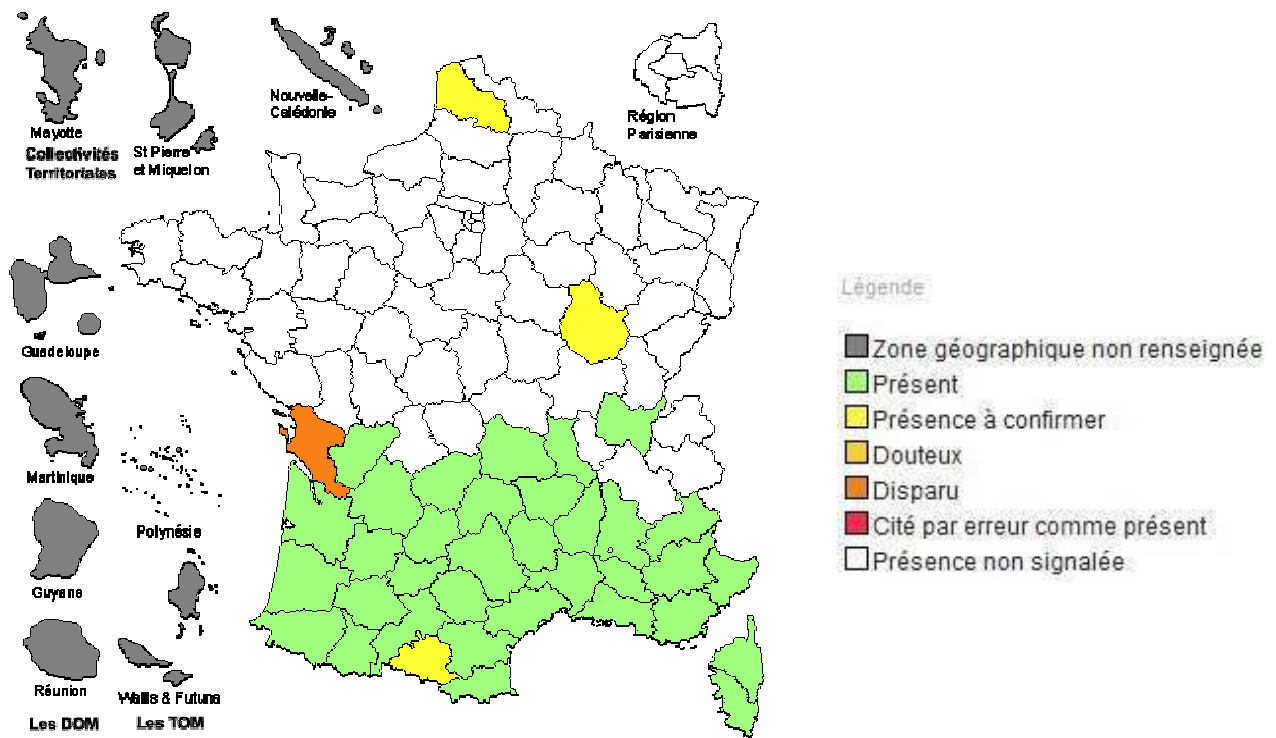
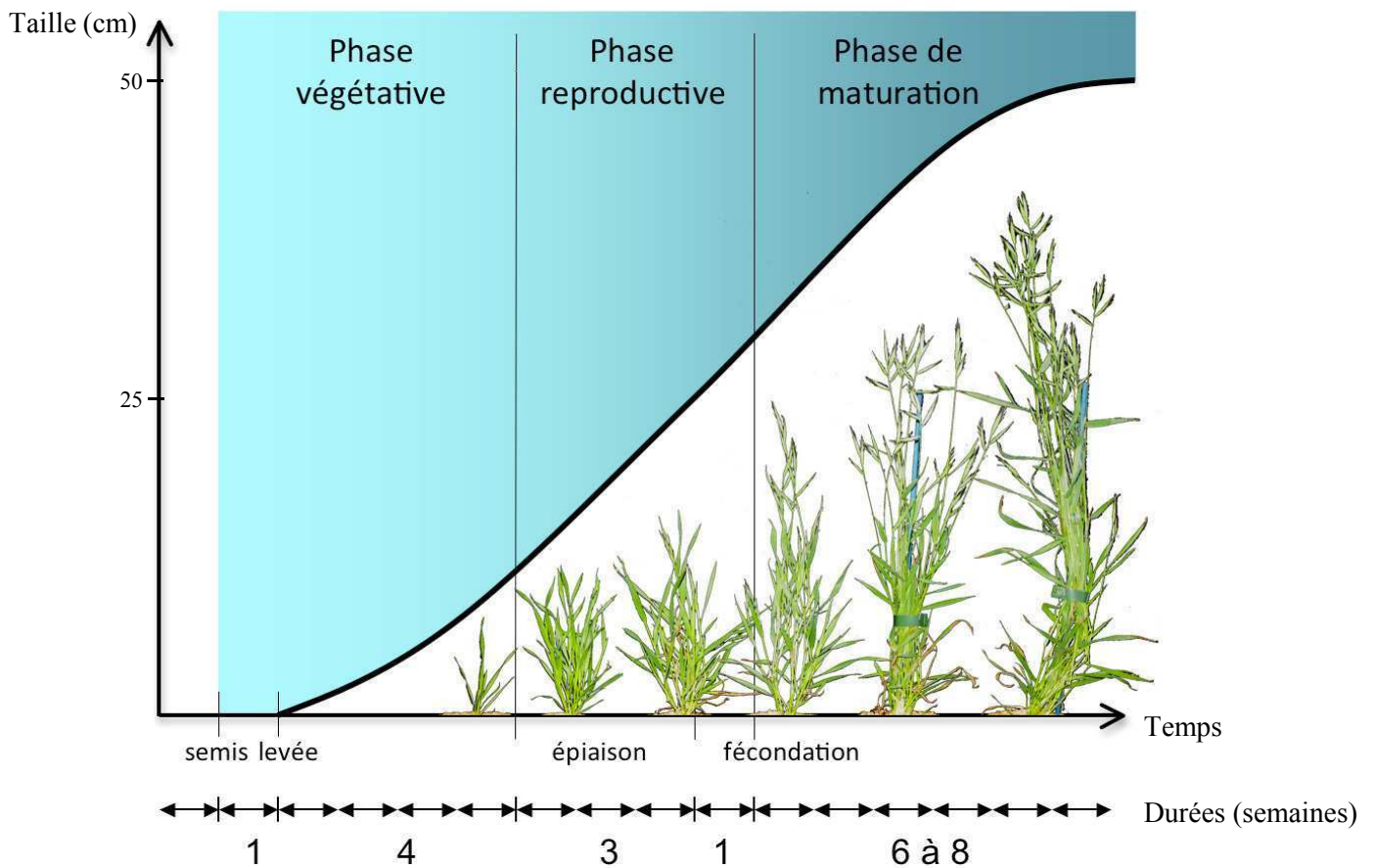


Figure 9. Répartition naturelle de *Brachypodium distachyon* en France. (<http://www.tela-botanica.org/eflore/BDNFF/4.02/mn/10075/chorologie>)

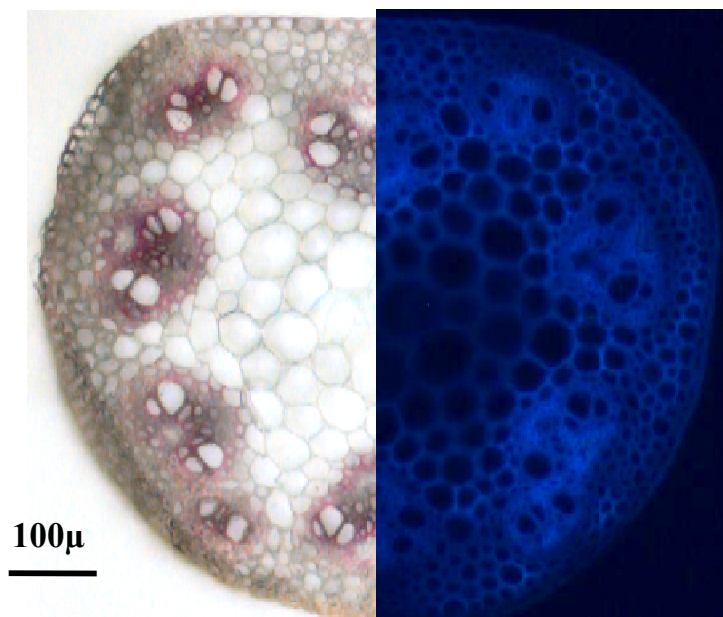
continent Américain et même en France (Fig. 8). Ainsi, *Brachypodium distachyon* est présente dans la moitié Sud de la France métropolitaine et est classée comme espèce protégée dans la région Poitou-Charentes (par l'arrêté du 19 avril 1988 relatif à la liste des espèces végétales protégées) (Fig. 9).

### 2.3. Croissance et développement

*Brachypodium* a un cycle de vie court, de 12 à 18 semaines dans les conditions expérimentales de Versailles: dans les chambres climatisées, avec une vernalisation de 4 semaines en jours courts (8 h de jour à 21°C et 18°C la nuit), puis un passage en jours longs (16 h de jour à 21°C et 18°C la nuit), 16 semaines sont nécessaires de graine à graine, avec un très bon rendement en graines (de 50 à 200 par plante). Il est cependant possible d'accélérer son développement en conditions de jours très longs (20 h de jour) ou de jours continus, mais les plantes obtenues sont alors plus petites et ont un rendement en graines nettement inférieur. On distingue d'abord la phase végétative, d'une durée de 7 semaines dans les conditions standards décrites précédemment, correspondant à la mise en place des organes végétatifs que sont les racines, les feuilles et les tiges constituées de l'alternance de nœuds et entre-nœuds. La tige (nœuds et entre-nœuds) d'une Poacée se développe de bas en haut, et les premiers entre-nœuds sont alors les plus âgés. Cependant, le développement de chaque entre-nœud étant acropétal, la partie la plus basse des entre-nœuds est la plus jeune, alors que la partie la plus haute est la plus mature et par conséquent plus lignifiée. A chaque entre-nœud correspond une feuille. La feuille est décrite en trois parties : la gaine entourant la partie basale de l'entre-nœud correspondant, la ligule et le limbe. La phase reproductrice débute lorsque le dernier nœud du brin maître produit le pédoncule et les épis. Les autres brins se développent ensuite de la même manière. Chaque inflorescence porte un épi de 3 à 7 épillets contenant chacun une dizaine de fleurs hermaphrodites qui se développeront en caryopses



**Figure 10.** Schéma du développement de *Brachypodium distachyon* dans les conditions de croissance de Versailles



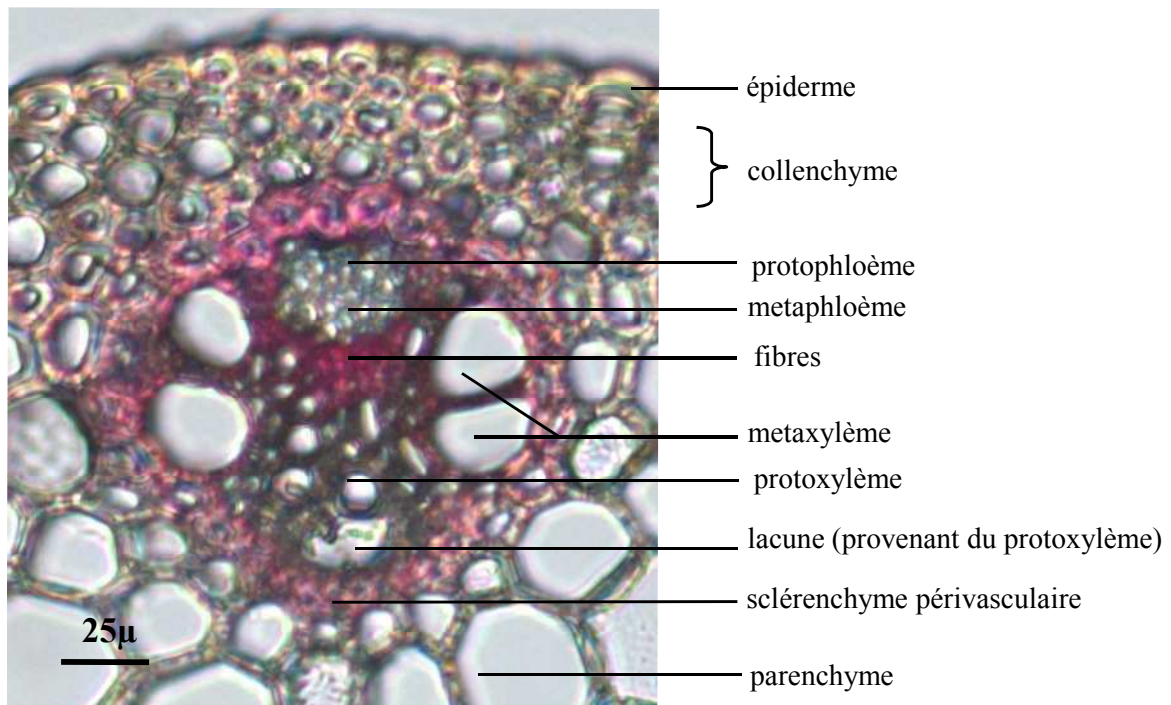
**Figure 11.** Coupe transversale d'un entre-nœud de *Brachypodium* colorée au phlorogucinol (tissus lignifiés en rose) et sous U.V (autofluorescence de la lignine à 530 nm).

après fécondation. La maturation des grains nécessite environ 6 semaines (Fig.10). La reproduction autogame de *Brachypodium* facilite la maintenance de lignées pures.

Les entre-nœuds sont pleins au début de leur développement, puis se creusent en leur centre. Le parenchyme en constitue le tissu majoritaire et les tissus vasculaires sont regroupés en faisceaux, répartis circulairement et régulièrement (Fig. 11). Ces faisceaux sont les tissus les plus lignifiés de la plante. Ils sont entourés d'une gaine de sclérenchyme lignifié, formée d'une seule couche de cellules dans les jeunes faisceaux et de deux ou trois couches de cellules dans les faisceaux plus âgés. Chaque faisceau est orienté et comprend différents types cellulaires (Fig. 12) avec, de l'intérieur vers l'extérieur : des lacunes (formées par la rupture des cellules de protoxylème), le protoxylème, le métaxylème (de chaque côté de fibres), le métaphloème et enfin le protophloème. Le protoxylème est constitué de trachéides qui, à maturité, seront des cellules mortes uniquement faites de leur paroi. La lignification de ces cellules se fait au niveau d'épaississements en spirale, ce qui leur permet de s'allonger pendant une partie de la croissance de la plante. Les lacunes se forment lorsque leur paroi épaissie ne leur permet plus de s'allonger. La circulation de la sève brute s'y fait par capillarité. Le métaxylème se forme ensuite. Les parois de ses cellules sont entièrement lignifiées et les parois transversales sont ponctuées (Shane, McCully et al. 2000). Les fibres sont des cellules mortes, très allongées, au diamètre réduit et dont les parois sont épaisses et entièrement lignifiées.

A maturité et dans les conditions de culture décrites précédemment, une plante mesure environ 30 centimètres de hauteur et nécessite une surface minimale de 0,0078 m<sup>2</sup> (1 plante par pot de 10 cm de diamètre, ce qui permet de faire pousser jusqu'à 130 plantes par mètre carré. Ces conditions de culture relativement simples en font un modèle cultivable dans des installations expérimentales habituelles.





**Figure 12.** Faisceau vasculaire observé au microscope à épifluorescence sur coupe transversale de jeune entre-nœud de *Brachypodium* coloré au phloroglucinol (coloration rose des tissus lignifiés).

## 2.4. Outils de travail disponibles

Le séquençage 4X du génome de l'accession Bd21 a été achevé en 2008 et le 8X en 2009, puis disponible et publié en 2010 par l'International Brachypodium Initiative (Vogel, Garvin et al. 2010). Ce génome a été annoté et peut être consulté sous forme d'ADNg, d'ADNc obtenus d'après le résultat de l'annotation, ou de séquences protéiques (provenant de la traduction *in silico* des ADNc) sur les sites suivants : <http://www.brachypodium.org/> et <http://mips.helmholtz-muenchen.de/plant/brachypodium/>. Ces sites contiennent également des collections de BAC (Hasterok, Marasek et al. 2006; Huo, Lazo et al. 2008), et d'EST (Vogel, Gu et al. 2006), ainsi que des outils bioinformatiques pour les étudier, tels que des BLAST pour comparer des séquences nucléotidiques ou protéiques, des programmes de recherche de domaines protéiques ou d'éléments génétiques (par BAC, Contigs ou noms). L'identification de marqueurs microsatellites Simple Sequence Repeat (SSR) a ensuite donné lieu à des cartes physiques et génétiques. Des puces Affymetrix® sont à présent disponibles, permettant des études d'expression à l'échelle du génome entier. Grâce aux nouvelles données sur le génome et son expression, d'autres outils bioinformatiques sont alors disponibles ou en cours de développement, tels que Bradinet (<http://aranet.mpimp-golm.mpg.de/bradinet/>) renseignant sur l'expression de gène en fonction des organes et des stades de croissance, ou Phytozome (<http://phytozome.net/>) permettant entre autres des études de synténie entre Brachypodium et d'autres espèces.

Brachypodium est une espèce naturellement autogame mais des techniques de croisements ont pu être mises au point, notamment par le laboratoire de Garvin en 2009 (consultable sur <http://www.docstoc.com/docs/41553551/Illustrated-Guide-to-Crossing-Brachypodium>). Pour ces croisements, il s'agit d'attendre le stade où les anthères ont commencé à se développer suffisamment pour être manipulables, mais bien avant maturité de la gynécée, puis d'ouvrir la fleur en soulevant (voire en coupant) les glumelles mais pas la lemme, d'en retirer les sacs



polliniques et de refermer la lemme avec soin. Il faut ensuite attendre que l'ovule soit à maturité (entre quelques heures et une journée) et déposer sur le stigmate des sacs polliniques ou des grains de pollen frais. Ces croisements sont délicats à effectuer et leurs taux de succès ne sont pas encore proches de 100%. Cependant ils ont permis de grandes avancées de la cartographie de *Brachypodium* : on peut à présent effectuer des back-cross ou des complémentations alléliques à partir des diverses accessions disponibles sous forme de lignées pures ou de différents mutants, permettant de rechercher des QTL ou de vérifier la ségrégation des traits phénotypiques ou encore de valider le rôle de certaines mutations.

La transformation de *Brachypodium* a été possible dès 2001 par bombardement biolistique (Draper 2001). La transformation par *Agrobacterium tumefaciens* a été mise au point quelques années plus tard (Vain, Worland et al. 2008; Vogel and Hill 2008): elle se fait *in vitro* sur cals embryogènes et permet de limiter le nombre de mutations dans le génome. La transformation génétique est un outil indispensable à l'utilisation d'un organisme comme modèle car elle donne accès à de nombreux outils tels que des collections de mutants par insertion d'ADN-T ou des lignées transgéniques exprimant des gènes ou séquences d'intérêt sous contrôle de promoteurs choisis. Il existe également à présent des collections de mutants par irradiation ou par mutagenèse chimique. C'est en particulier le cas de la collection obtenue par mutagenèse à l'éthyle méthanesulfonate (EMS) et l'azoture de sodium ( $\text{NaN}_3$ ) grâce à la collaboration entre le centre INRA de Versailles et l'URGV d'Evry. Cette collection est exploitée à la fois par des techniques de génétique directe (phénotypage) et de génétique reverse par TILLING (McCallum, Comai et al. 2000) et a été utilisée dans les travaux décrits dans la suite de ce manuscrit.

Pour ces raisons, le développement de *Brachypodium distachyon* en tant que système modèle et l'étude de ses nombreuses accessions naturelles, particulièrement les accessions



Bd21 et Bd21-3, permettront d'établir des connaissances et des ressources pour améliorer les Poacées de grande culture pour des usages traditionnels (alimentation animale et humaine) ou plus innovants (utilisation pour la fabrication d'énergies renouvelables).

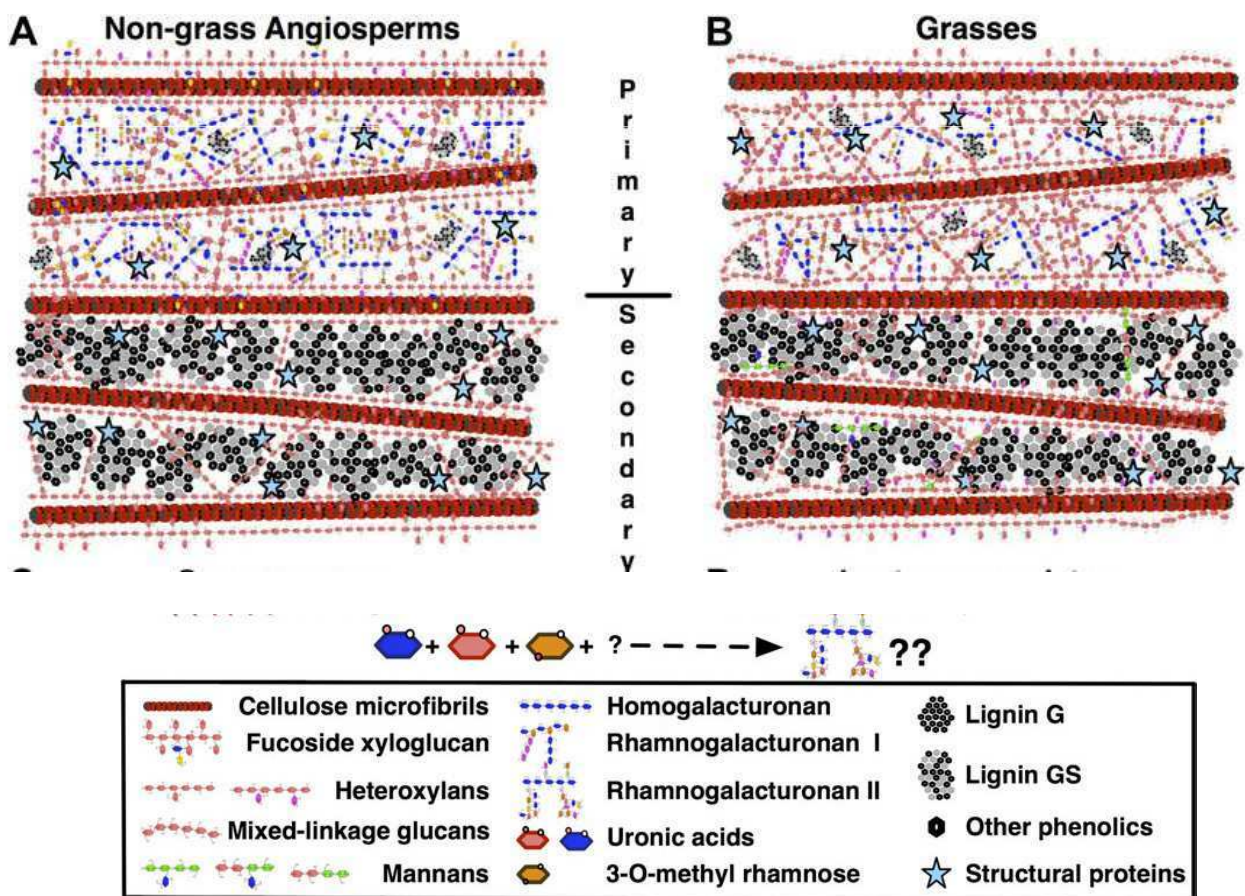
### 3. Organisation de la paroi lignocellulosique végétale

Les cellules végétales sont entourées d'une paroi rigide mais cependant dynamique (tant qu'elle n'est pas lignifiée), la paroi lignocellulosique. Celle-ci évolue au cours de la croissance et de la différenciation des cellules, grâce au dépôt de strates pariétales successives. Les parois ont un rôle structural majeur : une épaisseur de 0,1 à 10  $\mu\text{m}$  leur permet d'assurer un support mécanique, renforcé par la pression de turgescence qu'elles permettent de maintenir en donnant leur forme aux cellules et aux végétaux. Leur composition influant sur leur rigidité et leur imperméabilité, elles peuvent également constituer une barrière physique aux insectes et aux micro-organismes phytopathogènes, ou même à l'eau. Ainsi, elles contribuent au contrôle de l'évapotranspiration au niveau des organes exposés au milieu extérieur, ou à la redirection de l'eau de l'apoplasme au symplasme dans les racines grâce aux cadres de subérine, ou encore au transport des sèves dans les tissus conducteurs. De plus, la paroi cellulaire végétale joue un rôle dans la signalisation entre cellules adjacentes (Pilling and Hofte 2003) assurant un contrôle du développement de la plante. La communication entre les cellules a également lieu par le biais de ponctuations de la paroi, les plasmodesmes, permettant la continuité entre le cytoplasme des cellules adjacentes. Les parois cellulaires sont très diverses au sein du monde végétal, mais il existe pour toutes des structures et une architecture communes, et sont formées selon les mêmes principes. Cependant, selon les plantes ou selon les tissus, elles peuvent présenter des différences significatives de composition. Ainsi, des différences notables d'organisation et composition

**Tableau 3.** Comparaison des compositions pariétales des parois primaires et secondaires des monocotylédones et dicotylédones (Vogel 2008 Unique aspects of grass cell walls, corrigé d'après source: Ishii 1997).

	Primary wall		Secondary wall	
	Grass	Dicot	Grass	Dicot
Cellulose	20–30 <sup>b,c</sup>	15–30 <sup>c,d,e</sup>	35–45 <sup>c,f</sup>	45–50 <sup>c</sup>
Hemicelluloses				
Xylans	20–40 <sup>d</sup>	5 <sup>c</sup>	40–50 <sup>c,g</sup>	20–30 <sup>c,g</sup>
MLG	10–30 <sup>d</sup>	Absent	Minor	Absent
XyG	1–5 <sup>c,d,g</sup>	20–25 <sup>g</sup>	Minor	Minor
Mannans and glucomannans	Minor	5–10 <sup>d</sup>	Minor	3–5 <sup>g</sup>
Pectins	5 <sup>c</sup>	20–35 <sup>d</sup>	0.1 <sup>c</sup>	0.1 <sup>c</sup>
Structural proteins	1 <sup>d</sup>	10 <sup>d,e</sup>	Minor	Minor
Phenolics				
Ferulic acid and $\rho$ -coumaric acid	1–5 <sup>c,d</sup>	Minor (except order Caryophyllales)	0.5–1.5 <sup>c</sup>	Minor (except order Caryophyllales)
Lignin	Minor	Minor	7–10 <sup>c</sup>	20 <sup>c</sup>
Silica			5–15 <sup>c</sup>	Variable

<sup>a</sup> Numbers in this table were taken from several sources to provide rough approximations of generalized cell wall composition from typical dicots and grasses. Some of the numbers are averages or ranges based on multiple sources.



**Figure 13.** Comparaison des compositions pariétales simplifiées en deux dimensions des parois primaires et secondaires des monocotylédones et dicotylédones (Sakar 2009) Les microfibrilles de cellulose sont présentes dans les deux groupes. (A) Angiospermes dicotylédones (B) Poacées.

pariétales existent entre les parois des poacées (ou graminées ou grass en Anglais) et celles des autres angiospermes. Celles-ci sont récapitulées Tableau 3 et Fig. 13 et concernent principalement les hémicelluloses : les parois des poacées sont caractérisées par l'abondance d'arabinoxylanes estérifiés par l'acide férulique servant de site d'arrimage aux lignines, ainsi que par leur faible teneur en pectines. De plus, il existe des différences importantes de composition entre strates pariétales (parois primaires et parois secondaires), la principale étant l'abondance des lignines dans les parois secondaires des tissus de soutien et conducteurs. Lors d'une division cellulaire, la paroi est synthétisée *de novo* et trois couches principales sont formées, la lamelle moyenne, la paroi primaire et la paroi secondaire (qui peut être divisée en strates S1, S2 et S3) (Fig.14).

### 3.1. La lamelle moyenne

La première couche pariétale synthétisée est la lamelle moyenne commune aux deux nouvelles cellules. Elle provient de la plaque cellulaire formée lors de la division d'une cellule mère en cellules filles. Dans les tissus pas ou peu lignifiés, cette lamelle moyenne est riche en pectines et en protéines. Dans les tissus fortement lignifiés, la lamelle moyenne contient la plus forte concentration en lignines (50% ou plus). La lamelle moyenne joue un rôle fondamental dans la cohésion entre les cellules.

### 3.2. La paroi primaire

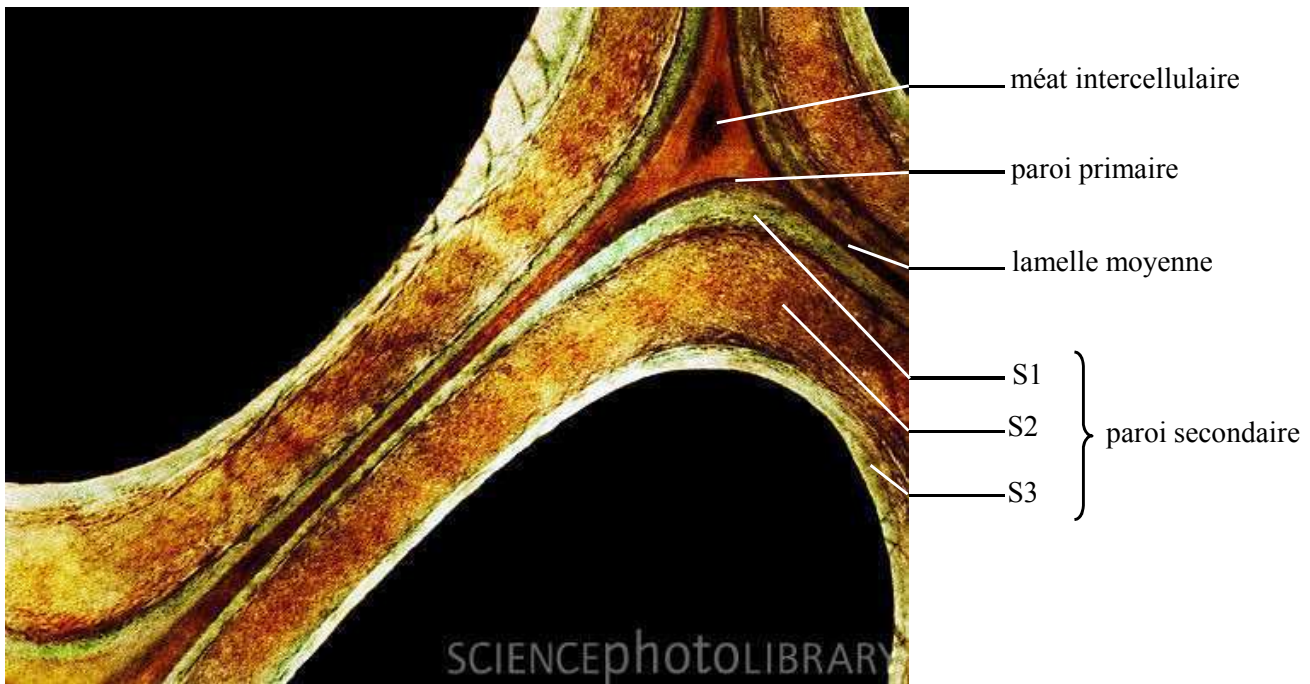
Sa formation débute pendant la division cellulaire, où elle vient s'apposer de chaque côté de la lamelle moyenne. Elle sera ensuite renforcée pendant la phase de croissance cellulaire (Sánchez-Rodríguez, Rubio-Somoza et al. 2010). Dans les cellules non lignifiées, cette paroi primaire est hydrophile, élastique et ne mesure au maximum que quelques



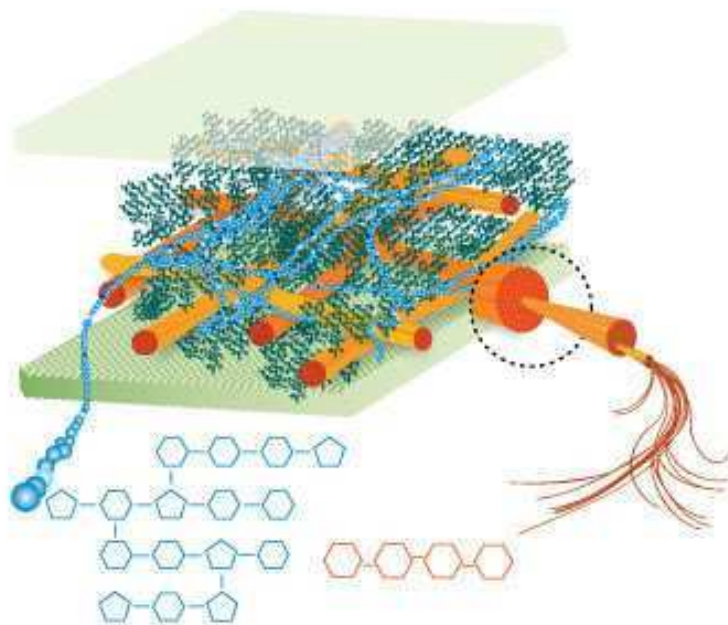


dixièmes de micromètres d'épaisseur. Elle est constituée de microfibrilles de cellulose enchâssées dans une matrice amorphe de pectines, d'hémicelluloses et de glycoprotéines. Sa plasticité permet à la cellule de continuer sa croissance en longueur. Il existe deux types de paroi primaire, se différenciant par les proportions des composés de leur matrice. Le type I, caractéristique des dicotylédones mais trouvé également chez certaines monocotylédones (les monocotylédones non commélinoïdes, telles que celles de l'ordre des Liliales contenant la famille des liliaceae, ou encore l'ordre des Asparagales contenant la famille des orchidaceae), contient principalement de la cellulose et des hémicelluloses en quantités équivalentes, puis des pectines et des protéines structurales. Ses hémicelluloses sont en majorité des xyloglucanes (Yokoyama and Nishitani 2004). Le type II, dont le constituant majoritaire est également la cellulose, contient une plus faible quantité de pectines, de protéines et de xyloglucanes. En revanche, ce type de paroi contient une quantité plus importante d'hétéroxylyanes (les glucuronoarabinoxylanes dont les substituants arabinose sont partiellement estérifiés par l'acide férulique et, dans une moindre mesure, l'acide *p*-coumarique) et, particulièrement chez les Poacées, de (1-3;1-4)- $\beta$ -D-glucanes (MLG pour mixed linkage glucan) qui sont des chaînes non branchées et non substituées de résidus (1,3)- et (1,4)- $\beta$ -glucosyl associées aux cellules en croissance (Yokoyama and Nishitani 2004; Burton and Fincher 2009; Fincher 2009). Les parois de type II sont spécifiques des monocotylédones commélinoïdes, telles que les céréales, les carex ou les joncs (Smith and Harris 1999).

Toutes les cellules végétales possèdent une paroi primaire, à l'exception de certaines cellules reproductrices. Les cellules de parenchyme ou de collenchyme sont des exemples de cellules ne possédant généralement qu'une paroi primaire, alors que pour d'autres types cellulaires celle-ci est complétée par une paroi secondaire.



**Figure 14.** Paroi végétale de *T. canadensis* observée au Microscope Electronique à Transmission et colorée, sur laquelle se distinguent les différentes couches de la paroi cellulaire (source: © NAS/Photo Researchers/OKAPIA)



**Figure 15.** Modélisation en trois dimensions de l'organisation de la paroi végétale (source: ceres.net). En Bleu les hémicelluloses, en orange la cellulose et en vert les lignines et acides phénoliques

### 3.3. La paroi secondaire

Dans le cas des tissus de soutien et des tissus conducteurs, tels que le sclérenchyme, les fibres ou les éléments vasculaires, plusieurs couches successives de paroi secondaire viennent s'apposer à la paroi primaire après l'arrêt de la croissance en longueur et leur différenciation. Dans la majorité des cas, trois couches sont formées : S1, la première à partir de la paroi primaire, S2, et enfin S3, la couche la plus interne à la cellule. Ces couches se différencient par l'orientation des microfibrilles de cellulose qui les constituent. La couche S2, généralement la plus épaisse, détermine la résistance mécanique des cellules. Certaines espèces végétales peuvent contenir un plus grand nombre de couches, jusqu'à six dans le cas des fibres de bambou (Gritsch and Murphy 2005). L'ensemble de la paroi secondaire, plus épaisse que la paroi primaire, peut atteindre jusqu'à 10  $\mu\text{m}$  et est principalement constituée de cellulose, d'hémicelluloses et, dans la majorité des cas, de lignines (Taylor and Turner 2007). Les 2 types d'acides liés aux parois de poacées peuvent se dimériser soit par réaction photochimique ce qui conduit à des dimères de type cyclobutane, soit, dans le cas de l'acide férulique, par oxydation peroxydasique conduisant aux diférulates. Dans les tissus lignifiés, la paroi est inextensible et imperméable. En terme de quantité, la paroi secondaire représente la biomasse la plus importante produite par les plantes terrestres, et le plus grand réservoir de carbone organique de la biosphère. Une représentation tridimensionnelle simplifiée de paroi secondaire est donnée Fig. 15.



## 4. Constituants de la paroi végétale

### 4.1. Protéines pariétales

Il existe de nombreux types de protéines pariétales : celles agissant sur ou interagissant avec les polysaccharides, des oxydo-réductases, des protéases, des protéines structurales, des protéines impliquées dans la signalisation et les réponses aux stress et aux pathogènes (<http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/>).

Des protéines dont les fonctions sont encore inconnues ont été identifiées suggérant de potentiels rôles additionnels (Showalter 2001; Jamet, Albenne et al. 2008; Albenne, Canut et al. 2009).

#### 4.1.1. Protéines structurales

Ce sont les protéines majoritaires des parois cellulaires végétales. Elles contribuent à leur architecture en contrôlant leur assemblage, leur expansion, leur niveau d'hydratation et de perméabilité et peuvent servir de site nucléation pour la lignification ou de source de molécules-signal.

Elles ont été classées en quatre groupes selon leur composition :

- Les glycoprotéines riches en hydroxyproline (HRGPs) chez les plantes autres que les graminées, dont les rôles restent à élucider mais incluant les extensines, participant à l'élongation cellulaire en permettant l'allongement de la paroi. Les orthologues des expansines chez les graminées sont des glycoprotéines riches en hydroxyproline et en thréonine (THRGPs), ces protéines pouvant être réticulées par des ponts entre tyrosine, ce qui bloque l'extension des parois. Les extensines sont à la fois considérées comme protéines structurales et comme enzymes pariétales (Carpita 1996).
- Les protéines riches en proline (PRPs), dont certaines ont été décrites chez le maïs, impliquées dans le développement en modulant l'architecture de la paroi, dans la



maintenance de l'intégrité de la paroi et potentiellement dans la lignification (Cassab and Varner 1988).

- Les protéines riches en arabinogalactanes (AGPs), ou protéoglycanes de la matrice extracellulaire, impliquées dans de nombreux processus liés à la croissance et au développement des cellules (Showalter 2001).
- Les protéines riches en glycine (GRPs), dont tous les rôles n'ont pas encore été élucidés mais qui sont localisées dans les tissus vasculaires auxquels elles apportent vraisemblablement leur élasticité au cours de leur développement. Certaines ont été identifiées chez le riz et l'orge. Chez les graminées, beaucoup de ces protéines ont vraisemblablement des fonctions cytosoliques suggérées par l'absence de peptide signal (Carpita 1996).

#### 4.1.2. Enzymes de modification pariétale

Les parois cellulaires contiennent de nombreuses enzymes contribuant à leur plasticité par la modification de leurs composants polysaccharidiques. Leur rôle est particulièrement important lors de la croissance et la différenciation cellulaire. Ce sont en majorité des hydrolases qui peuvent favoriser le relâchement puis l'extension de la paroi directement en hydrolysant les liaisons entre microfibrilles de cellulose (cellulases), ou indirectement en réduisant la taille et la viscosité des polymères de la matrice (pectine méthylestérases, pectine acétylestérases, féruloyl estérases) permettant l'action d'autres protéines telles que les expansines. Elles pourraient également avoir d'autres rôles, notamment dans la défense contre les pathogènes, la signalisation ou l'utilisation des polysaccharides pour des besoins métaboliques ou énergétiques. Cette classe regroupe des endo-glycanases ( $\beta$ -galactosidase), des exoglycanases ( $\beta$ -glucanase ou  $\beta$ -glucosidase) et des polygalacturonases (Cosgrove 1999). On y trouve notamment les  $\beta$ -glucosidases qui sont potentiellement impliquées dans la déglucosylation des formes glucosylées des monolignols avant leur incorporation au polymère





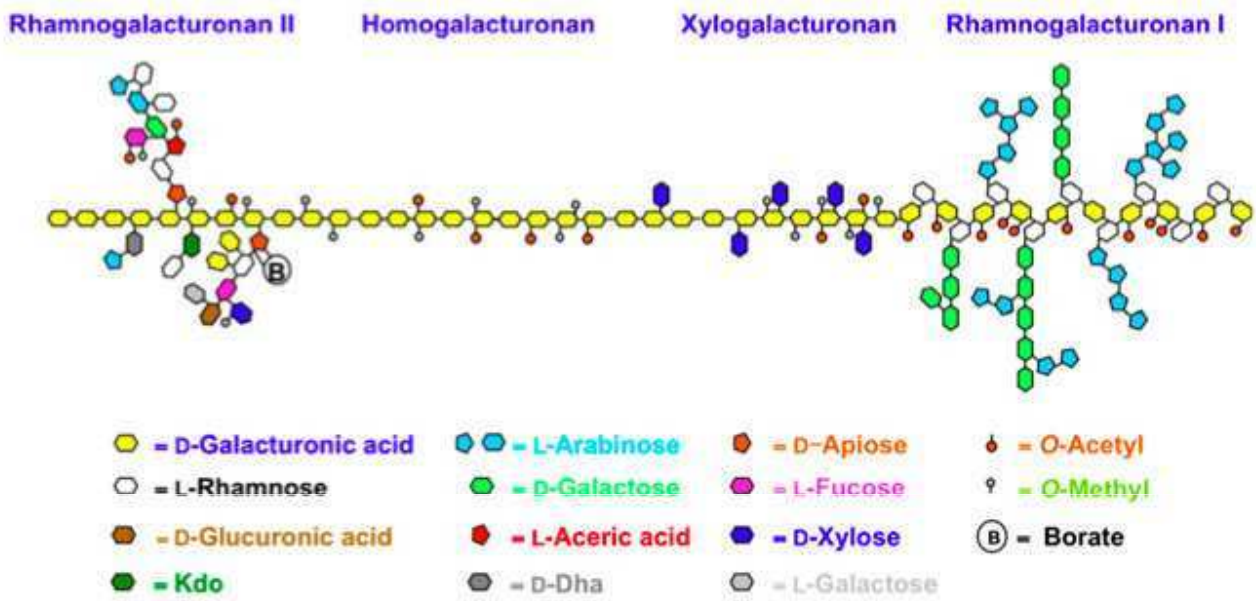
de lignine. Il existe également des protéines de dégradation spécifiques de certains constituants pariétaux qui ne sont ni des hydrolases ni des oxydases: la dégradation des composés pectiques est assurée par des pectate lyases, des pectine lyases et des rhamnogalacturonane I lyases et pourrait jouer un rôle dans la signalisation intercellulaire, particulièrement lors de l'infection par des pathogènes.

La polymérisation des monolignols en lignines au niveau des parois nécessite leur oxydation préalable afin que leur couplage puisse avoir lieu. Il a été montré *in vitro* que des peroxydases secrétées par les plantes ont la capacité d'effectuer cette étape. Il s'agit de peroxydases de classe III, impliquées dans de nombreux processus cellulaires. Les peroxydases constituent des familles multigéniques comportant de nombreux membres : soixante treize chez *Arabidopsis thaliana*, classées en six groupes (Welinder, Justesen et al. 2002), cent cinquante cinq chez le maïs, classées en sept groupes ou cent quarante neuf chez le riz (<http://peroxibase.toulouse.inra.fr>).

Cependant, d'autres protéines localisées dans les parois pourraient être impliquées dans ce processus : les laccases. Ce sont également des familles multigéniques contenant dix sept membres chez *Arabidopsis* (Sato, Wuli et al. 2001), vingt sept ont été annotées chez le maïs, vingt neuf chez le riz (<http://cellwall.genomics.purdue.edu>) et vingt neuf également chez *Brachypodium*.

#### 4.2. Silice

La silice, sous forme d'inclusions dans l'épiderme, le périderme et certaines cellules spécialisées des racines, rhizomes et des tiges, est particulièrement abondante dans les parois cellulaires des céréales. La présence de silice augmente la rigidité et la résistance mécanique



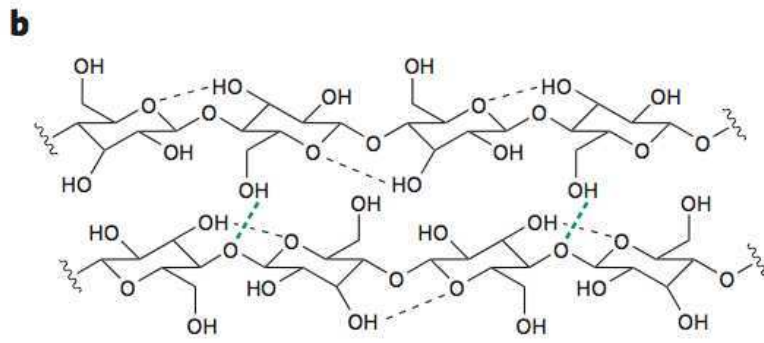
**Figure 16.** Structure schématique des pectines présentes dans les parois cellulaires végétales. (Harholt et al., 2010)

des parois, favorisant la croissance et la défense contre les stress biotiques et abiotiques (chimiques ou physiques). Deux mécanismes ont été suggérés : le premier étant que la silice associée à la cuticule agit en tant que barrière physique empêchant la pénétration des pathogènes, le second proposant que la silice sous forme soluble agirait en tant que modulateur de la résistance aux pathogènes (Ma and Yamaji 2008).

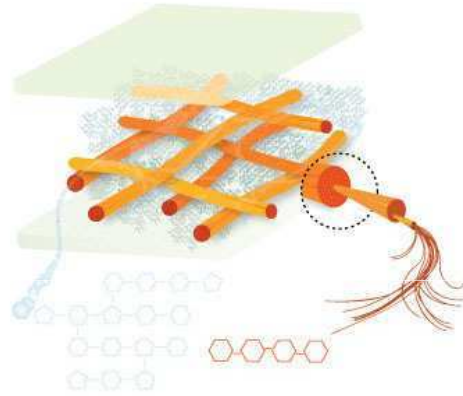
#### 4.3. Pectines

Les pectines sont présentes dans chacune des couches de la paroi cellulaire végétale, particulièrement abondantes dans les parois de cellules en cours de croissance ou de division, dans la lamelle moyenne et les coins cellulaires. Les parois primaires non lignifiées des dicotylédones en contiennent jusqu'à 35% et celles des céréales entre 2 et 10%. Les parois secondaires en contiennent une très faible proportion d'environ 0.1% (Vogel 2008). Les pectines sont des constituants structuraux importants de la paroi, et influencent leurs propriétés telles que leur porosité, leur charge de surface (surface charge), leur pH, et leur équilibre ionique, jouant un rôle dans les transports d'ions dans la paroi. Elles sont importantes pour la rigidité des parois (Ryden, Sugimoto-Shirasu et al. 2003). Elles jouent également un rôle dans les réponses de défense aux pathogènes, en servant d'éliciteurs déclenchant les cascades de signalisation lorsqu'elles sont dégradées (D'Ovidio, Mattei et al. 2004).

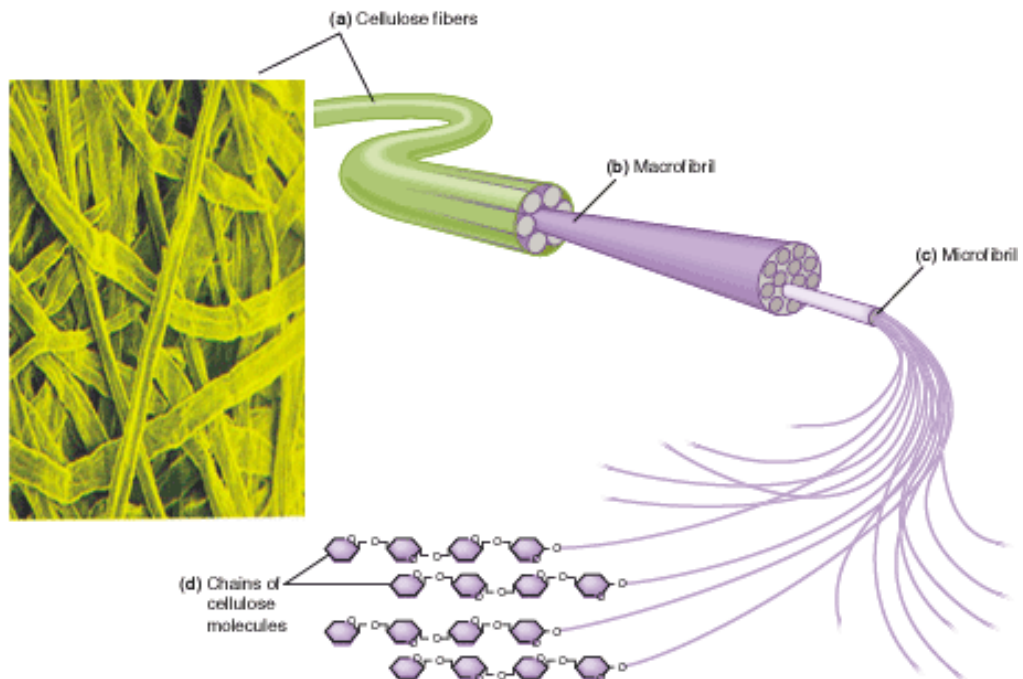
Les pectines sont des hétéro-polysaccharides majoritairement constituées d'acide galacturonique (Fig. 16). Le polysaccharide le plus abondant (65%) est l'homogalacturonane (HG), un homopolymère pouvant atteindre jusqu'à 100 acides galacturoniques liés en  $\alpha$ -1-4. D'autres pectines sont des HG substitués : Le rhamnogalacturonane II (RG-II) (10%), le xylogalacturonane (XGA), l'apigalacturonane (AP) et le rhamnogalacturonane I (RG-I) (20 à



**Figure 17.** Modélisation du polymère de cellulose dans la paroi (Endler et al., 2010)



**Figure 18.** Modélisation de la cellulose dans la paroi végétale (source: ceres.net)

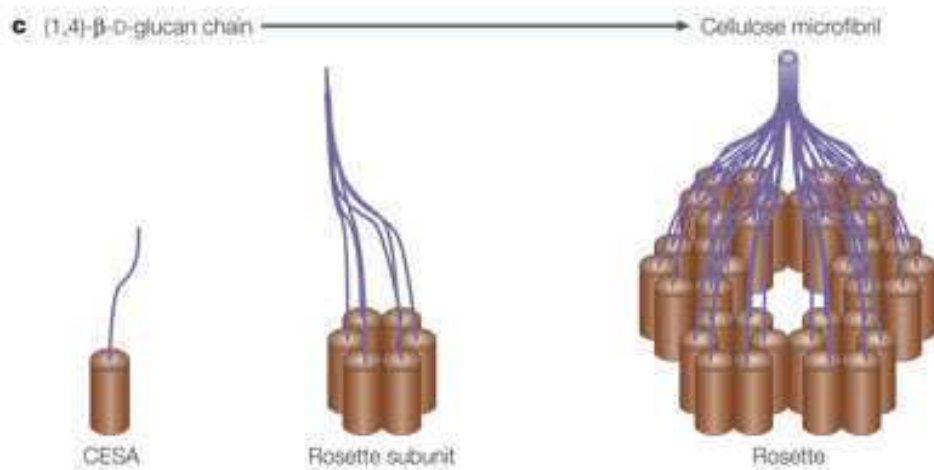


**Figure 19.** Structure réticulée de la cellulose de (<http://hrsbstaff.ednet.ns.ca/>)

35%). Les RG-II sont particulièrement importants pour la croissance et le développement, et des modifications même mineures peuvent engendrer des défauts de croissance. Les pectines sont synthétisées dans l'appareil de Golgi puis transportées à la paroi dans des vésicules membranaires. La biosynthèse des pectines est un procédé complexe nécessitant l'intervention de nombreuses glycosyltransférases, méthyltransférases et acétyltransférases. L'activité de ces protéines a été montrée *in vitro* mais seul un petit nombre de gènes codant ces enzymes ont été formellement identifiés *in planta* (Mohnen 2008); (Harholt, Suttangkakul et al. 2010).

#### 4.4. Cellulose

C'est le polymère commun à toutes les parois végétales. C'est également le constituant majoritaire de ces parois. Il s'agit d'un polymère linéaire de glucose liés en  $\beta$ 1-4, et dont l'unité de base est le cellobiose (Fig. 17). Une seule chaîne de cellulose peut compter jusqu'à 5000 unités de glucose selon la strate pariétale et l'origine botanique. Ces chaînes de cellulose (jusqu'à une soixantaine) sont assemblées en microfibrilles (Fig. 18 et 19) d'un diamètre final de 3 à 5 nm et longues de plusieurs micromètres (Cosgrove 2005). Le caractère cristallin mais également la solidité de cet assemblage est dû aux liaisons hydrogène établies entre chaînes adjacentes (Fig. 17). Dans la paroi primaire, les microfibrilles de cellulose sont fines et orientées aléatoirement, ce qui permet la croissance isotropique des cellules, c'est à dire dans toutes les directions nécessaires, bien que légèrement orientées parallèlement à l'axe de croissance. Cet axe est déterminé par l'ensemble des contraintes s'exerçant sur la cellule qui a alors une croissance anisotrope (Crowell, Gonneau et al. 2010). Dans chaque strate de paroi secondaire, les microfibrilles possèdent une orientation privilégiée, variant entre les différentes strates. Généralement, dans la strate S1 la plus proche de la paroi primaire, c'est-à-dire la couche la plus interne de la paroi secondaire, elles ont une orientation proche de l'axe



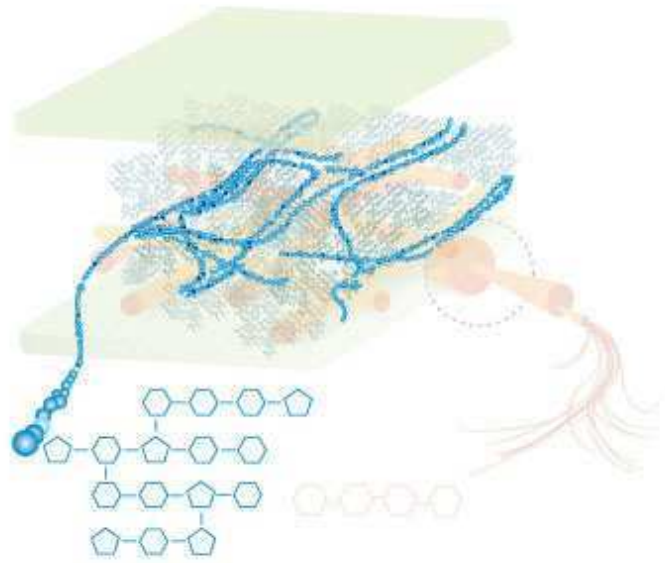
Copyright © 2005 Nature Publishing Group  
 Nature Reviews | Molecular Cell Biology

**Figure 20.** Modèle du complexe de synthèse de cellulose. Les CESA sont représentées par les cylindres marrons et les microfibrilles de cellulose en bleu. (Cosgrove 2005)

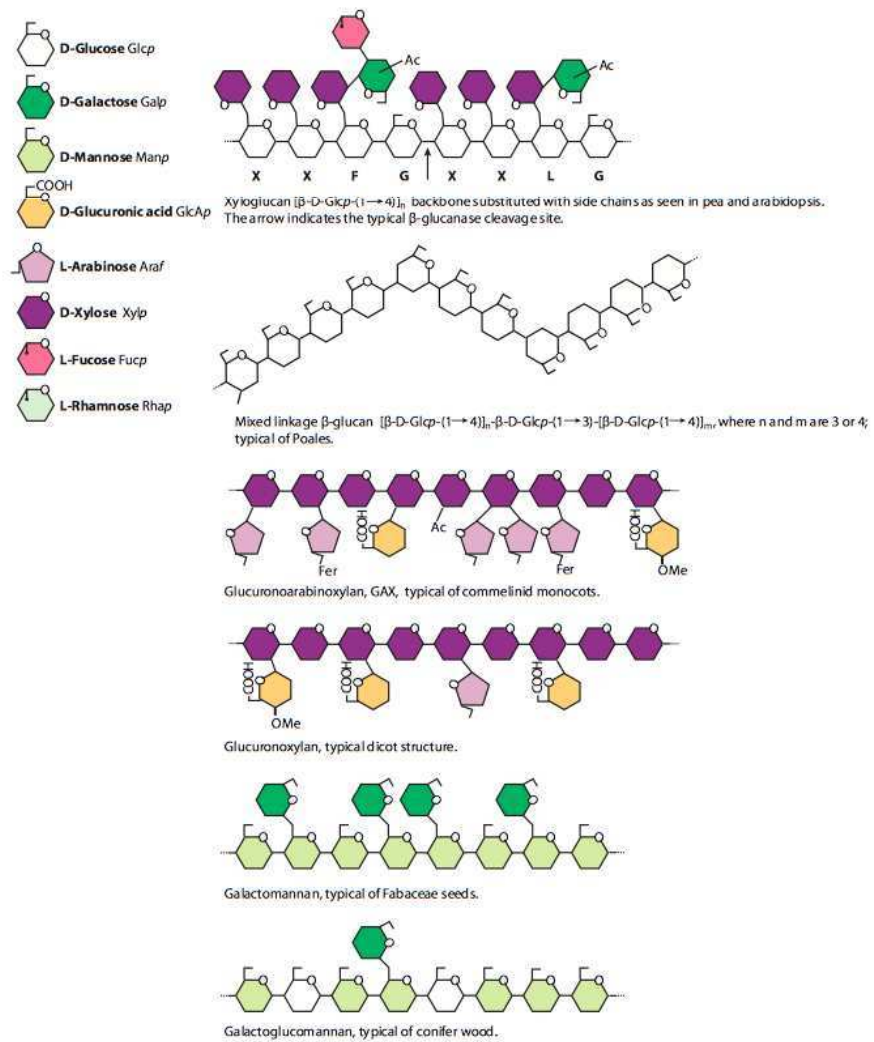
de la cellule. Plus les strates seront externes (S2 et S3), plus les microfibrilles de cellulose y prendront une orientation oblique par rapport à l'axe de la cellule. Leur orientation, qui dépend également de l'espèce et du tissu concerné (Déjardin, Laurans et al. 2010), est déterminée par la disposition de microtubules corticaux (Baskin 2001; Harris, Bulone et al. 2010). La cellulose peut constituer jusqu'à 45% de la paroi secondaire (en % de la matière sèche). Grâce à ces caractéristiques, la cellulose forme la trame fibrillaire des parois végétales et est à l'origine de leur résistance à la traction.

La cellulose est synthétisée au niveau de la membrane plasmique par des complexes de cellulose synthase organisés en six rosettes constituées de six celluloses synthases (CESA) (Fig. 20) Les sous-unités des complexes CESA, dont certaines sont peut être interverties, varient selon les espèces, les stades de développement et le type de paroi (primaire ou secondaire) (Gardiner, Taylor et al. 2003). Chaque complexe permet de synthétiser trente six chaînes de cellulose qui s'associeront alors en microfibrilles puis seront exportées vers la paroi à travers les pores membranaires formés au centre des complexes de CESA (Joshi, Bhandari et al. 2004). Ces microfibrilles peuvent s'associer en macrofibrilles, et ces dernières se regroupent pour former les fibres de cellulose. La cellulose contenue dans les parois végétales est le polymère le plus abondant dans la biosphère mais l'organisation des parois lignocellulosiques peut constituer un obstacle à son exploitation, notamment lorsque la lignine doit en être extraite, et ce, par des procédés polluants, nécessitant beaucoup d'apports d'énergie et coûteux. Ce sont les fibres de cellulose qui sont ciblées, par exemple pour la fabrication de pâte à papier par voie chimique, ou la production de biocarburant, et la biomasse recherchée pour ces applications est alors celle contenant d'importantes quantités de cellulose.





**Figure 21.** Modélisation de l'hémicellulose dans la paroi végétale (source: ceres.net)



**Figure 22.** Schéma des structures des hémicelluloses présentes dans les parois cellulaires végétales (Scheller et al., 2010)

#### 4.5. Hemicelluloses

Les hémicelluloses sont des hétéropolysaccharides constitués principalement de glucose, galactose, mannose, xylose, rhamnose, fucose et d'arabinose (Scheller and Ulvskov 2010) (Fig. 21 et 22). Ce sont des polysaccharides dont la chaîne principale est faite de sucres liés en  $\beta$ -(1-4) présents dans toutes les parois végétales, avec pour exception les liaisons  $\beta$ -(1-3,1-4) de MLG qui ne sont présentes que dans la famille des Poacées et quelques autres familles de l'ordre des Poales. Les hémicelluloses contiennent également de l'acide glucuronique. Les hémicelluloses sont synthétisées au niveau de l'appareil de Golgi puis exportées vers la paroi et peuvent contenir entre 300 et 5000 sucres. Les hémicelluloses sont très hydrophiles et peuvent former un gel lorsqu'elles sont hydratées. Elles se fixent aux microfibrilles de cellulose par des liaisons hydrogène. Dans la paroi primaire, les hémicelluloses sont très abondantes et permettent d'éviter que les microfibrilles de cellulose ne s'associent entre elles, jouant un rôle non négligeable dans leur élasticité (Pauly, Albersheim et al. 1999; Somerville, Bauer et al. 2004). Les hémicelluloses sont également liées aux lignines dans les parois lignifiées, mais par des liaisons covalentes directes entre sucres et unités de lignines et, dans le cas des Poacées également par des ponts féruliques. Ainsi, leur rôle biologique majeur est de renforcer les parois végétales par leurs interactions avec les autres composants.

Les hexoses (principalement le glucose) contenus dans les hémicelluloses peuvent être ciblés lors de la saccharification, notamment dans le cadre de la production de biocarburants. Cependant, les pentoses (majoritairement du xylose) qu'elles contiennent ne sont pas facilement fermentescibles (Yang and Wyman 2004; Jeoh, Ishizawa et al. 2007).

Il existe différents types d'hémicelluloses : leur structure varie selon les espèces, les tissus et les stades de développement. En particulier, les hémicelluloses des Poacées contiennent en majorité des glucuronoarabinoxylanes (GAX), constituées d'une chaîne principale de xyloses

**Tableau 4.** Exemples de concentrations d'acide férulique lié par des liaisons ester présentes dans les parois de différents organes ou tissus au sein de diverses espèces végétales. (Harris 2010)

Group	Species, family, order	Organ or tissue	Concentration of ferulic acid ( $\mu\text{g g}^{-1}$ cell walls)	Reference
Commelinid monocotyledons	Barley ( <i>Hordeum vulgare</i> ), Poaceae, Poales	Straw (stem)	3,510	Hartley (1987)
	Maize ( <i>Zea mays</i> ), Poaceae, Poales	Grain	20,660	Hartley (1987)
	Perennial ryegrass ( <i>Lolium perenne</i> ), Poaceae, Poales	Shoot	6,030	Hartley (1987)
		Leaf mesophyll cells	5,740–6,290	Harris et al. (1980)
	Chinese water chestnut ( <i>Eleocharis dulcis</i> ), Cyperaceae, Poales	Peeled corm	7,217	Parr et al. (1996)
	Chufa ( <i>Cyperus esculentus</i> ), Cyperaceae, Poales	Peeled tuber	4,492	Parker et al. (2000)
Non-commelinid monocotyledons	Asparagus ( <i>Asparagus officinalis</i> ), Asparagaceae, Asparagales	Stem (shoot) (green)	78–96	Rodríguez et al. (2005)
	Onion ( <i>Allium cepa</i> ), Alliaceae, Asparagales	Bulb	7	Ng et al. (2000)
'Core' Caryophyllales (eudicotyledons)	Beetroot ( <i>Beta vulgaris</i> ), Amaranthaceae, Caryophyllales	Root	6,939	Waldron et al. (1997)
	Sugarbeet ( <i>Beta vulgaris</i> ), Amaranthaceae, Caryophyllales	Root	4,592	Waldron et al. (1997)
Other eudicotyledons				
Rosids	Alfalfa ( <i>Medicago sativa</i> ), Fabaceae, Fabales	Stem	20	Hartley (1987)
		Stem	50	Chen et al. (2006)
Rosids	<i>Arabidopsis thaliana</i> , Brassicaceae, Brassicales	Stem	27	Derikvand et al. (2008)
Rosids	Black cottonwood ( <i>Populus trichocarpa</i> ), Salicaceae, Malpighiales	Leaf	78	Gou et al. (2008)
		Stem	78–252	
		Wood (secondary xylem)	89	
Asterids	Carrot ( <i>Daucus carota</i> ), Apiaceae, Apiales	Root	30	Parr et al. (1997)
		Root	16	Kang et al. (2008)

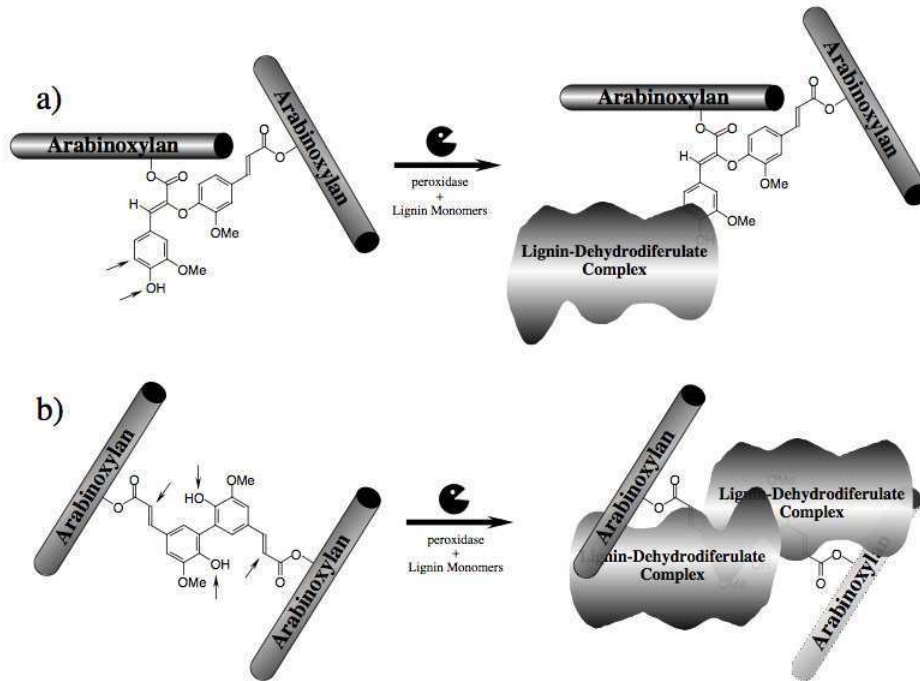
<sup>a</sup> The plants examined were not genetically manipulated, subjected to stress or grown in tissue culture

liés *en*  $\beta$  (1-4), substitués par des arabinoses et plus rarement par des acides glucuroniques liés *en*  $\alpha$ (1-2). Leurs substituants arabinose sont partiellement estérifiés par l'acide férulique et, dans une moindre mesure, l'acide *p*-coumarique. Ces hémicelluloses contiennent également des (1,3)-(1,4)- $\beta$ -D-glucanes (MLG, Mixed Linkage Glucans) qui sont propres aux parois de type II, donc des parois des Poacées. Les MLG sont des homopolymères non ramifiés et non substitués de glucose, constitués d'une série de plusieurs unités (1,4)- $\beta$ -D-glucose, entrecoupées par une unité liée *en* (1,3)- $\beta$ -D-glucose et associées aux cellules en croissance (Urbanowicz, Rayon et al. 2004; Yokoyama and Nishitani 2004; Burton and Fincher 2009; Fincher 2009).

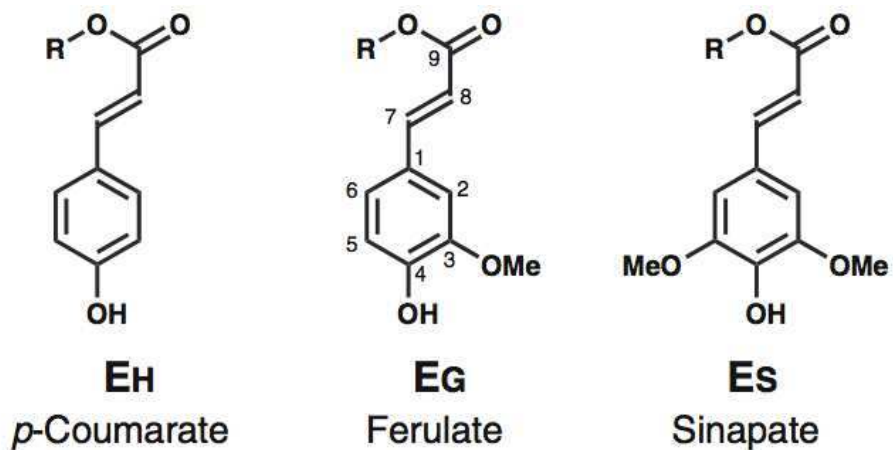
#### 4.6. Acides hydroxycinnamiques

Les acides hydroxycinnamiques sont représentés par l'acide para-coumarique (*p*CA), l'acide férulique (FA) et ses oligomères (déhydrodimères en particulier). Le FA et le *p*CA sont des intermédiaires de la voie de biosynthèse des monolignols. Cependant ils sont intégrés dans les parois des monocotylédones commélinoides (Tableau 4). Dans une moindre mesure, on trouve l'acide férulique estérifié aux pectines de la paroi primaire de certaines dicotylédones de l'ordre des Caryophyllales. Les céréales au métabolisme en C4 (maïs, sorgho, canne à sucre...) contenant plus d'esters *p*-coumariques que celles en C3, il serait prévisible d'en trouver de plus faibles quantités chez *Brachypodium* que chez le maïs par exemple (Harris and Trethewey 2010).

Chez les céréales, l'acide férulique sous forme de dimères (diférulate) est lié aux glucuronoarabinoxylanes de la paroi par des liaisons ester, participant au renforcement des parois et servant de points d'ancrage à la polymérisation et la réticulation des lignines en se liant aux lignines par liaison éther ou carbone-carbone (Ralph, Grabber et al. 1995) (Fig. 23). Il se dépose précocement, principalement pendant la mise en place de la paroi secondaire et



**Figure 23.** Dimères de férulates liés aux polysaccharides (chaines d'arabinoxylanes) de la paroi pouvant incorporer le polymère de lignine. (Source: Ralph 2006, US dairy forage Research center)



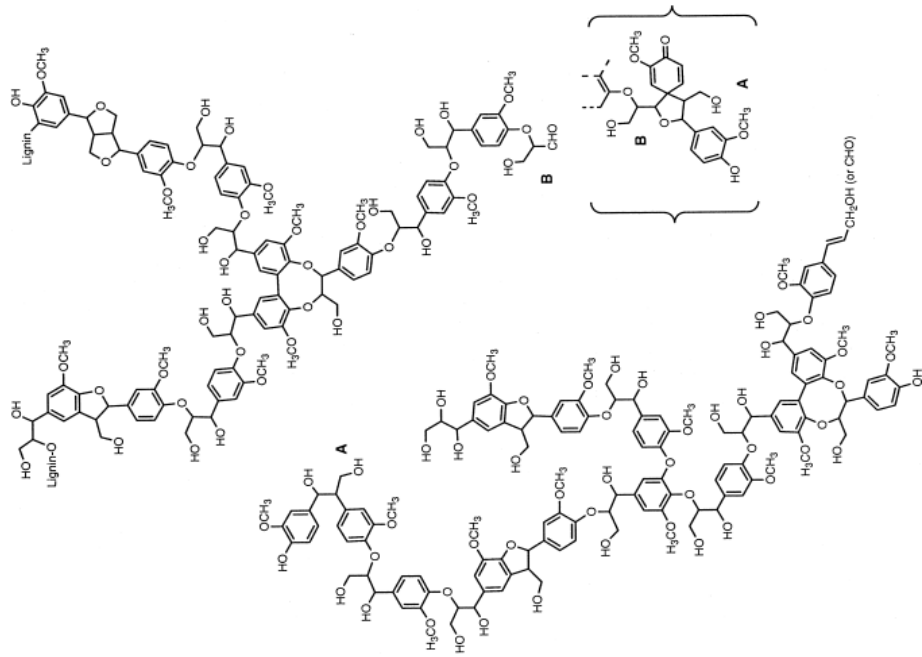
**Figure 24.** Structures des acide hydroxycinnamiques (Ralph, 2010).

au début de l'étape de lignification (He and Terashima 1991). L'acide férulique est l'acide hydroxycinnamique majoritaire des parois non lignifiées. On en trouve des quantités allant de 3,5 mg/g chez l'orge à 20,7 mg/g de paroi chez le maïs (Tableau 4). Les monomères de férulate sont liés au polymère de lignine comme les autres monomères, par des liaisons éther ou C-C après leur oxydation. Il a été montré que la réduction de la quantité de diférulate ancrant les lignines aux polysaccharides de la paroi permet d'augmenter significativement les rendements de saccharification de ces parois (Grabber 2005).

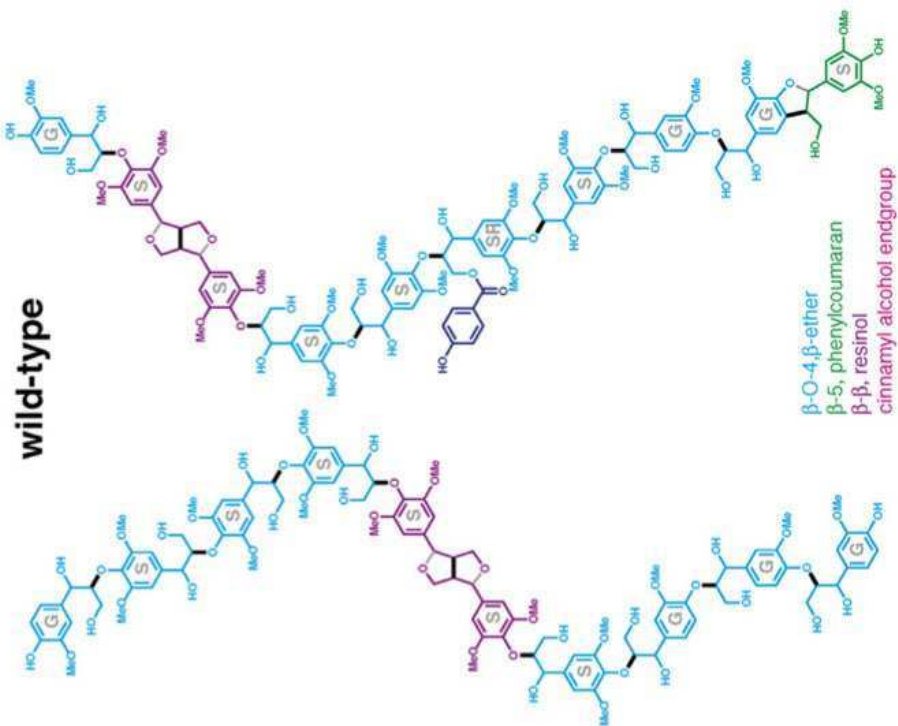
Si une faible partie estérifie les arabinoxylyanes, le *p*CA se lie principalement par liaisons ester sur les unités S et, chez le maïs, est un indicateur de leur dépôt dans la paroi (Grabber, Ralph et al. 2004). L'alcool sinapylique serait acylé par *p*CA et le sinapyl *p*-coumarate serait intégré au polymère de lignine (Lu and Ralph 2002). La présence de *p*CA a une influence sur la polymérisation des unités S, l'organisation spatiale des lignines, et leur capacité à se lier à la paroi. Il peut représenter jusqu'à 20% du contenu en lignines chez le maïs et le sorgho (Morrison III and Archibald 1998; Hatfield, Ralph et al. 1999; Ralph, Lundquist et al. 2004) mais a tendance à être en quantités plus importantes chez les plantes en C4 que chez celles en C3 (Ford, Elliott et al. 1987; Cherney, Mertens et al. 1990; Grabber and Jung 1991). Il se dépose de façon continue tout au long de la lignification (He and Terashima 1991) et est considéré comme un marqueur de la déposition de lignine chez les graminées (Musel, Schindler et al. 1997; Morrison III and Archibald 1998; Vailhé, Provan et al. 2000).

D'autres acides hydroxycinnamiques peuvent être présents dans la paroi, mais en plus faible quantité : l'acide sinapique, l'acide 5-hydroxyférulique et l'acide cafféique (Yamamoto, Lewis et al. 1987) (Fig. 24).

Les réseaux de composés phénoliques sont considérés comme assurant un rôle structural dans les parois des céréales après la fin de la croissance des cellules (Iiyama, Lam et al. 1990).



**Figure 25.** Structure d'un polymère de lignine de conifère illustrant la structure branchée du polymère (Brunow, 2005)



**Figure 26.** Structure d'un polymère de lignine de peuplier représentant les unités G et S (indiquées en gris) et quelques types de liaisons par lesquelles elles peuvent être liées. (Vanholme et al. 2005)

Cependant, la réticulation rend certains domaines polysaccharidiques inaccessibles aux enzymes de dégradation pariétale et peut inhiber l'accès à d'autres domaines.

## 4.7. Lignines

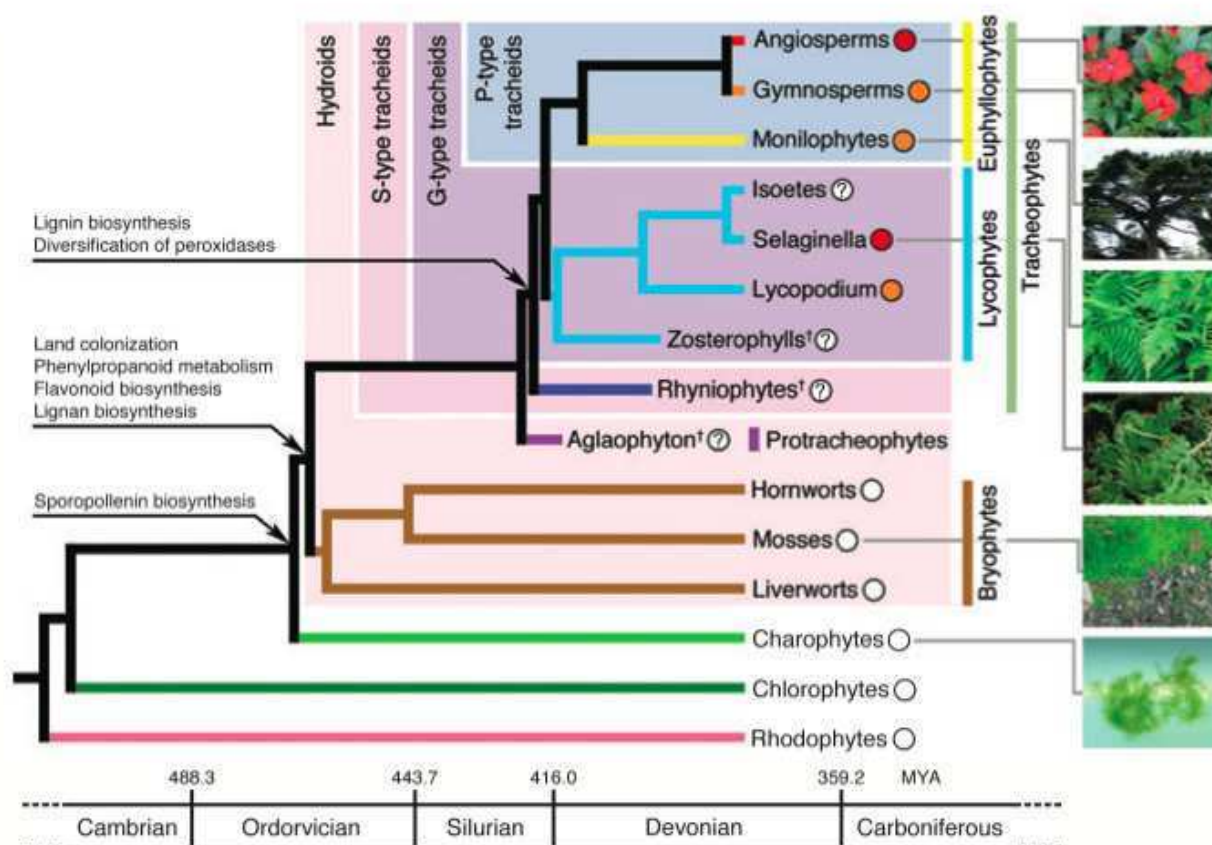
### 4.7.1. Lignification

Dans la plupart des cellules possédant une paroi secondaire, la lignification coïncide avec sa mise en place et a lieu dans l'intégralité des différentes couches de la paroi. Cette paroi contenant alors en majorité de la cellulose et de la lignine sera appelée paroi lignocellulosique. La lignification se fait de l'extérieur vers l'intérieur, c'est-à-dire en commençant par la lamelle moyenne, par les angles de la cellule, puis la paroi primaire, et pour terminer, les couches de paroi secondaire (Donaldson 2001). Par conséquent, la plus forte concentration en lignines est dans la lamelle moyenne mais la plus grande quantité relative est dans la paroi secondaire (Fergus and Goring 1970). Cependant, tous les types cellulaires ne contiennent pas de paroi secondaire. Chez les céréales, et chez *Brachypodium* en particulier, les tissus les plus lignifiés sont les cellules du xylème, les fibres, le sclérenchyme péri-vasculaire, le sclérenchyme périphérique et une partie du parenchyme périphérique situé dans les entre-nœuds (Fig.12). Des défauts de lignification peuvent être associés à des phénotypes particuliers, tels que des modifications de couleur des tissus lignifiés, des retards et des défauts de croissance, la présence de vaisseaux collapsés dans les faisceaux conducteurs, ou encore des anomalies de port érigé.

### 4.7.2. Description des lignines

La lignine (Fig. 25 et 26) est un terme générique représentant un ensemble de polymères phénoliques complexes issus principalement de la polymérisation de 3 alcools





**Figure 27.** Apparition de la lignine coïncidant avec celle des trachéophytes (Jing-Ke Wang and Clint Chapple, the origin and evolution of lignin biosynthesis, New phytol 2010)

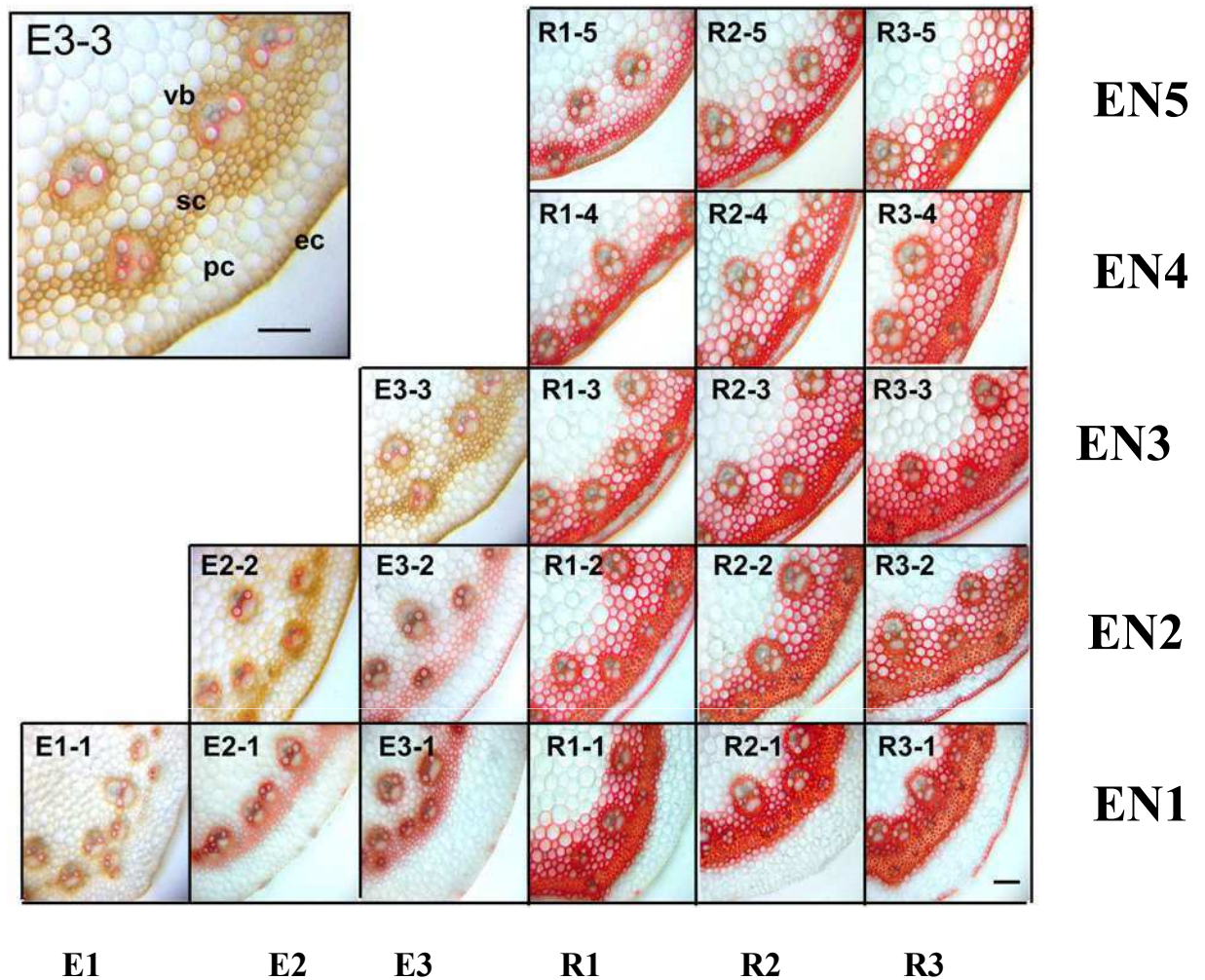
**Tableau 5.** Rendement en lignine et fréquence des monomère H, G et S des parois de différentes espèces (Barrière et al., 2007). Les pourcentages de chaque monomère H, G et S sont indicatifs des espèces considérées. La présence des unités H est bien plus importante chez les céréales que chez les dicotylédones.

	<b>Klason lignin %</b>	<b>Monomer yield</b>	<b>H</b>	<b>G</b>	<b>S</b>
Wheat straw	19.5	1190	4.6	43.2	52.2
Triticale straw	19.8	1610	3.0	42.0	55.0
Rye straw	18.8	1670	2.0	44.0	54.0
Maize stalk	14.8	930	2.3	37.4	60.3
<i>Arabidopsis</i> stem	17.3	1022	0.7	72.3	27.1
Tobacco stem	19.1	1868	0.3	51.7	48.1
Poplar wood	23.4	2310	0.3	36.7	63.0

hydroxycinnamiques (ou monolignols), l'alcool *p*-coumarylique, l'alcool coniférylique et l'alcool sinapylique, précurseurs respectifs des unités *p*-hydroxyphényles (H), guaiacyles (G) et syringyles (S) des lignines. D'autres composés phénoliques peuvent être incorporés dans une moindre proportion aux polymères lignines, tels que le coniféraldéhyde, le sinapaldéhyde ou l'acide férulique.

Les lignines sont des polymères amorphes, c'est-à-dire qu'ils ne prennent pas de forme particulière prédéfinie mais occupent l'espace laissé libre par les autres constituants de la paroi. C'est cette occupation de l'espace, combiné à la structure des lignines qui donne sa rigidité et son imperméabilité à la paroi lignifiée. Ces propriétés permettent de définir la forme des cellules qui ne changera plus sous l'influence de la pression de turgescence. La paroi lignifiée a un rôle structural majeur permettant le maintien érigé de la plante et la résistance aux stress mécaniques (notamment la compression), et donnant la résistance nécessaire aux vaisseaux pour la conduction de la sève brute par les forces liées à l'évapotranspiration. Les lignines peuvent également être déposées en réponse à des stress biotiques ou abiotiques et limitent la biodégradabilité des parois.

La lignine est très abondante sur la planète, c'est la seconde source de carbone organique, juste après la cellulose. La lignine est apparue avec les trachéophytes, plantes vasculaires à port érigé (Fig. 27) et a probablement joué un rôle déterminant dans l'adaptation des plantes au milieu terrestre. Cependant, des phénylpropanoïdes solubles ont été retrouvés chez les bryophytes montrant que la mise en place de ce qui est devenu l'organisation métabolique de la biosynthèse des monolignols a eu lieu plus tôt au cours de l'évolution, probablement lors de la colonisation du milieu terrestre par les plantes primitives où la nécessité de la protection contre les rayons UV s'est imposée (Lowry, Lee et al. 1980; Wang, Tang et al. 2011). Des traces de lignines ont par ailleurs été découvertes chez certaines algues (Martone, Estevez et al. 2009) et des données phylogénétiques suggèrent une convergence évolutive de la



**Figure 28.** Observation de l'évolution de la lignification sur différents organes et à différents stades de croissance de l'ivraie par coloration au phloroglucinol de coupes transversales d'entrenœuds (EN). E1 à E3: trois étapes du stade d'élongation; R1 à R3: trois étapes du stade reproductif. Le second chiffre (1 à 5) correspond à la position de l'entre nœud observé, 1 décrivant l'EN le plus bas et 5 le dernier EN (Tu et al., 2010).

biosynthèse des lignines entre les algues rouges corallines et les plantes vasculaires. Depuis, les lignines ont suivi l'évolution et se sont complexifiées: les gymnospermes, apparues il y a 350 Ma ont une lignine principalement constituée d'unités G, avec quelques traces d'unités H. Les angiospermes dicotylédones, apparues il y a environ 250 Ma, ont une lignine constituée d'unités G et S en quantités équivalentes, et des traces de H. Les graminées, apparues il y a environ 70 Ma possèdent les trois unités H, S et G, les deux dernières représentant entre 85% et 96% du total. Le tableau 5 illustre la variabilité de la fréquence des unités H, G ou S chez quelques angiospermes (Barrière, Riboulet et al. 2007). Les proportions des différents types d'unités dans les lignines dépendent donc des espèces considérées, mais également des tissus ou des stades de développement. A titre d'exemple, chez les monocotylédones, les parois des cellules des faisceaux vasculaires contiennent principalement des lignines G, et les parois et tissus interfasciculaires contiennent en majorité un mélange de G et S (He and Terashima 1991). De plus, la lignification se fait d'abord avec des unités G et H, les unités S si elles sont présentes, ne se déposant qu'en fin de lignification (He and Terashima 1990; He and Terashima 1991). Les cellules jeunes seront moins lignifiées que les cellules âgées, de même que les tissus jeunes seront moins lignifiés que les tissus âgés. Ceci résulte en une hétérogénéité de la lignification à l'échelle de la plante, mais également à l'échelle d'un même organe ou tissu. Le développement acropétal de la tige de graminée aura donc pour conséquence que les premiers entre-nœuds (les plus bas) seront plus lignifiés que les derniers, et ce, jusqu'à maturation et lignification complète de la tige. Le développement basipétal des entre-nœuds fera au contraire qu'au sein de chacun de ces organes, la partie basse sera la moins lignifiée. En pratique, les premiers entre-nœuds étant très courts et se développant rapidement, ils atteignent en quelques semaines leur état final de lignification et ces différences ne sont plus observables (Fig. 28).

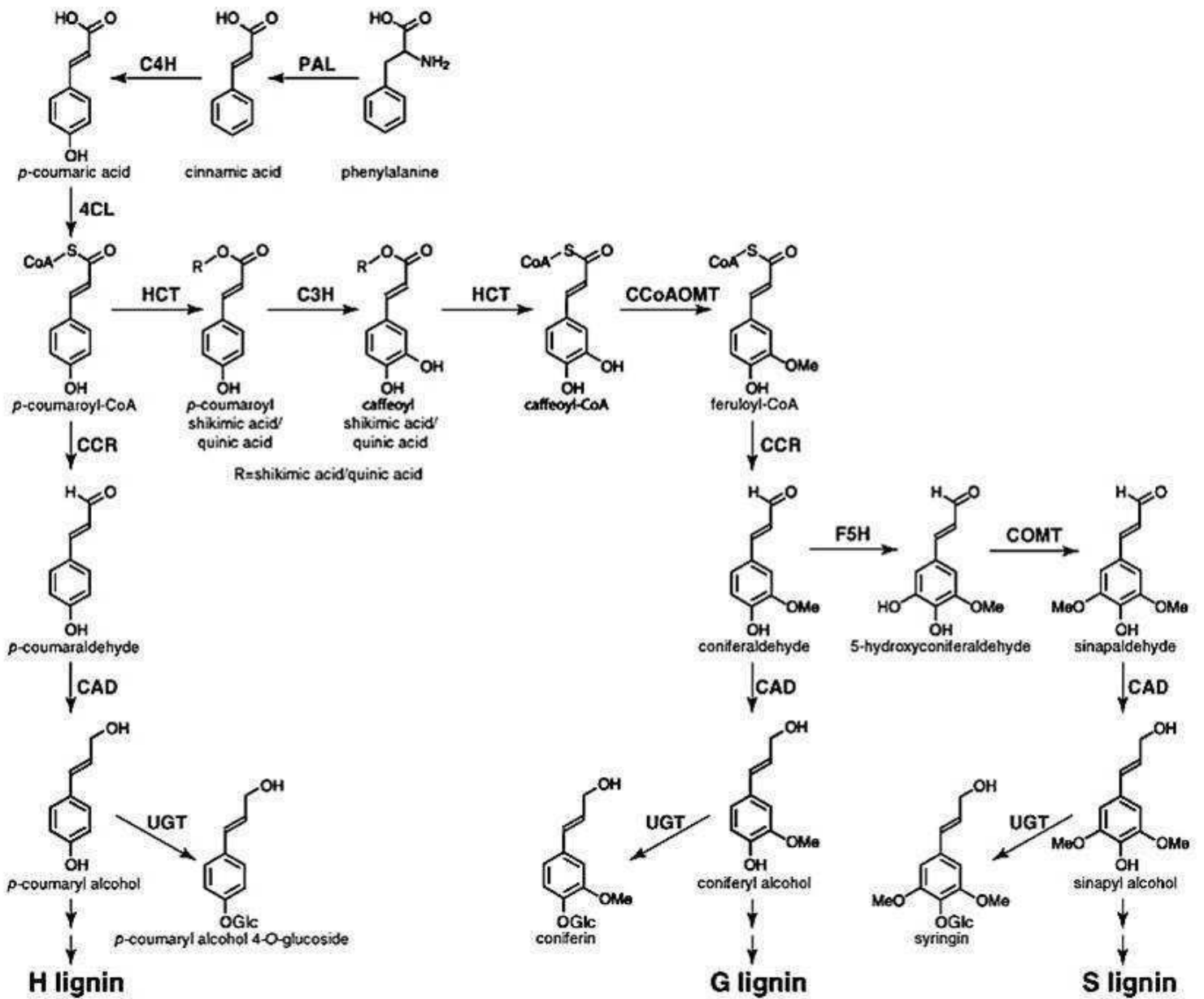


Figure 29. Voie principale de biosynthèse des monolignols (Vanholme et al., 2008)

Les lignines de graminées possèdent des particularités les démarquant des autres lignines décrites : on y trouve des unités S acylées par l'acide *p*-coumarique (Ralph, Hatfield et al. 1994; Grabber, Quideau et al. 1996) et des liaisons entre leurs unités G et les arabinoxylanes de la paroi assurées par de l'acide férulique (Jacquet, Pollet et al. 1995). Une autre particularité de ces lignines est la présence d'une importante quantité de groupes phénoliques libres (Lapierre 1993). La plupart des unités H sont en particulier présentes sous forme d'unités terminales à groupe phénolique libre. Cette richesse en groupes phénoliques libres rend compte de la bonne solubilité des lignines de graminées en milieu alcalin et à température ambiante : 50% des lignines de paille de monocotylédones peuvent être dissoutes à température ambiante dans la soude diluée, tandis que le même traitement appliqué à des parois de dicotylédones dissoudra moins de 10% des lignines.

#### 4.7.3. Biosynthèse des monolignols

Les trois monolignols principaux diffèrent par leur degré de méthylation de leur noyau aromatique : aucun pour l'alcool *p*-coumarylique, un pour l'alcool coniférylique et deux pour l'alcool sinapylique. La biosynthèse des monolignols (Fig. 29) se fait dans le cytosol des cellules en cours de lignification, à partir de la phénylalanine et, seulement chez les graminées, de la tyrosine. Ces acides aminés aromatiques sont produits dans les chloroplastes par la voie du shikimate puis exportés dans le cytosol. Les gènes de la voie de biosynthèse appartiennent à des familles multigéniques possédant des nombres de membres variables selon les espèces considérées (Tableau 6). Des mutants d'*Arabidopsis thaliana* dans chaque famille ont permis d'établir la voie décrite dans ce modèle. Cependant chez les graminées, espèces au génome plus vaste et plus complexe dû notamment aux nombreux événements de duplication au cours de leur évolution, un certain nombre de gènes ont été décrits comme impliqués dans la voie mais il reste encore des inconnues. Le travail sur une graminée modèle au génome plus réduit et séquencé telle que *Brachypodium* permet d'utiliser

**Tableau 6.** Nombre de membres annotés des familles de gènes de la voie de biosynthèse des lignines chez différentes espèces. Dans les cas où ces annotations ont été faites in-silico, un certain nombre des gènes annotés sont des gènes « like ».

	PAL	C4H	4CL	HCT	C3H	CCoAOMT	CCR	F5H	COMT	CAD
<i>A. thaliana</i>	4	1	4	1	3	7	2	2	1	9
<i>M. sativa</i>	4	1	10	6	1	4	18	3	26	21
<i>S. bicolor</i>	8	3	15	4	2	7	44	3	41	14
<i>O. sativa (Xu)</i>	14	4	16	9	1	11	55	3	38	5
<i>O. sativa (CW genomics)</i>	11		15	5		10	29		7	14
<i>P. trichocarpa</i>	6	3	22	7	4	7	40	4	35	21
<i>Z. mays</i>	10		11	38		6	19	8	3	7
<i>B. distachyon</i>	8	2	9			7	7	5	3	7

**Tableau 7.** Effets de modulations d'expression des gènes de biosynthèse des monolignols sur la quantité totale et des monomères H G et S des lignines. (Vanholme 2008)

Gene(s)	Total lignin	H	G	S	S/G	References
<i>PAL</i> ↓	↓	↓	↓	↓	↓/↑	[3,6,49]
<i>PAL</i> ↑	↑	n.a.	↑/No changes	↓/No changes	↓/No changes	[3]
<i>C4H</i> ↓	↓	↓	↓	↓	↓	[3,6]
<i>C4H</i> ↑	No changes	n.a.	No changes	No changes	No changes	[3]
<i>4CL</i> ↓	↓	↑	↓	↓	No changes	[3]
<i>HCT</i> ↓	↓	↑	↓	↓	↑	[6,13,15**]
<i>C3H</i> ↓	↓	↑	↓	↓	n.a.	[3,14*]
<i>CCoAOMT</i> ↓	↓	↑	↓	↓/No changes	↓/No changes/↑	[3,6,10]
<i>CCR</i> ↓	↓	↓	↓	↓	↓/↑	[3,7,17*]
<i>F5H</i> ↓	↓/No changes	n.a.	↑	↓	↓	[3,6]
<i>F5H</i> ↑	↓/No changes	n.a.	↓	↑	↑	[3]
<i>COMT</i> ↓	↓/No changes/↑	n.a.	↓/↑	↓	↓	[3,6,10]
<i>COMT</i> ↑	No changes	n.a.	No changes	No changes	No changes	[3]
<i>CAD</i> ↓	↓/No changes	n.a.	↑/No changes	↓/No changes	↓/No changes	[3,12]
<i>4CL</i> ↓ <i>F5H</i> ↓	↓	n.a.	n.a.	n.a.	↑	[3]
<i>CCoAOMT</i> ↓ <i>COMT</i> ↓	↓/No changes	n.a.	↓/No changes	↓	↓	[3,10]
<i>CCR</i> ↓ <i>COMT</i> ↓	↓	n.a.	n.a.	n.a.	↑	[3]
<i>CCR</i> ↓ <i>CAD</i> ↓	↓	n.a.	↓	↓	↑	[3]
<i>COMT</i> ↓ <i>CCR</i> ↓ <i>CAD</i> ↓	↓	n.a.	n.a.	n.a.	n.a.	[3]

l'annotation de gènes de ces familles pour effectuer des analyses fonctionnelles et identifier quels membres sont effectivement impliqués.

La voie décrite ici ne constitue que la voie principale de biosynthèse des monolignols et il a été montré que des réactions transversales et d'autres voies permettent également la biosynthèse des monolignols (Schoch, Morant et al. 2006) (Fig. 30). De plus, des variations de proportions de monomères incorporés, ainsi que l'incorporation de monomères atypiques peuvent avoir lieu, naturellement ou notamment en cas de perte d'expression de gènes de la voie de biosynthèse (Tableau 7).

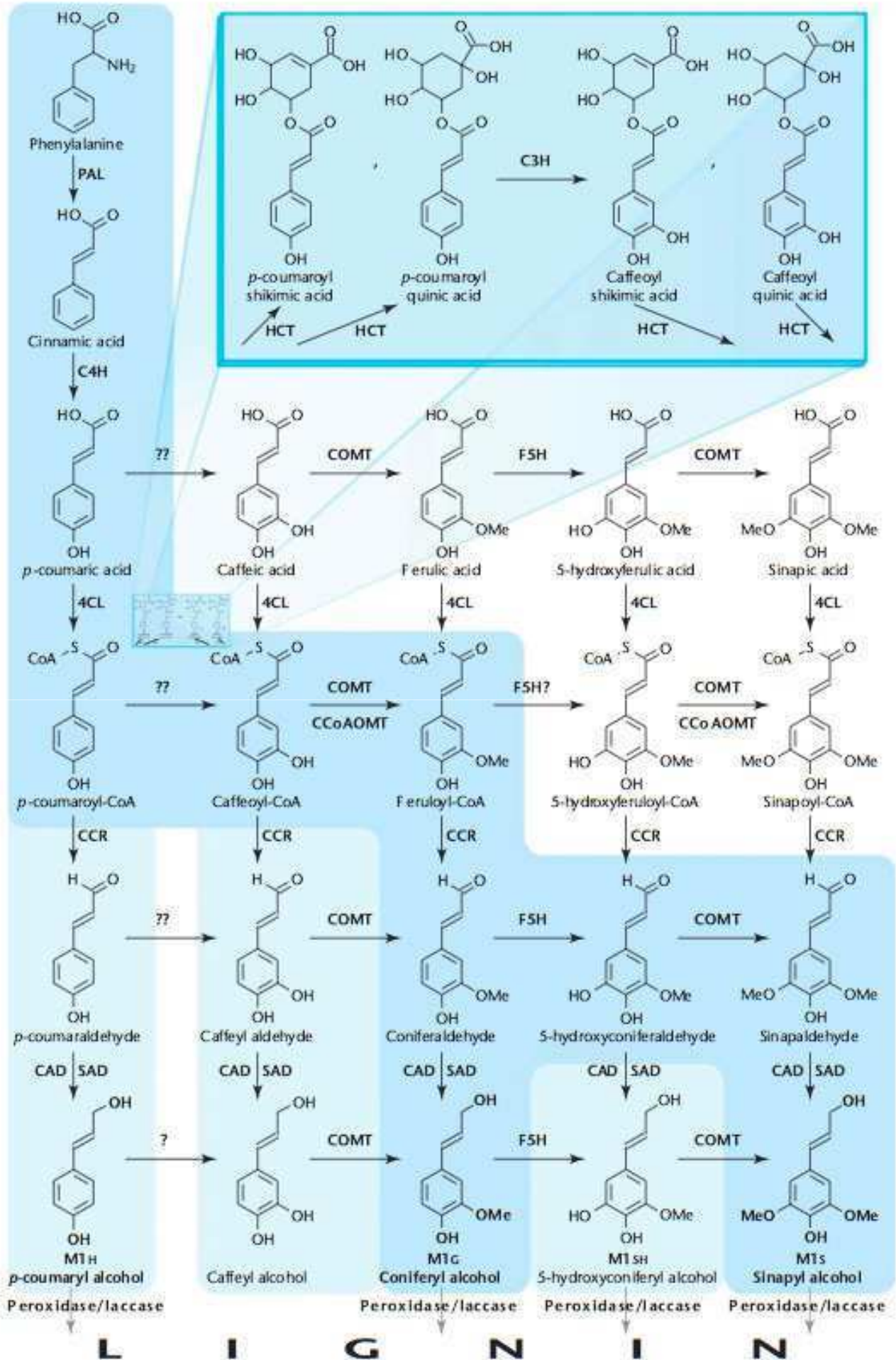
#### 4.7.3.1. Etapes communes de la voie des phénylpropanoïdes

La voie de biosynthèse des monolignols fait partie de celle des phénylpropanoïdes (Fig. 31). Elle commence par la désamination catalysée par la phénylalanine ammonia-lyase (PAL) de la chaîne latérale de la phénylalanine, ou de la tyrosine. Une PAL a été identifiée avec certitude et deux autres potentiellement pour leur implication dans cette voie chez le maïs (Rosler, Krekel et al. 1997). Elles possèdent également une activité tyrosine ammonia-lyase (TAL) catalysant la désamination de la tyrosine en acide *p*-coumarique (Rosler, Krekel et al. 1997). Quatre ont été identifiées chez *Arabidopsis thaliana* (Cochrane, Davin et al. 2004). Huit gènes *PAL* ont été annotés chez *Brachypodium*, mais sans certitude sur leur fonction.

Le cycle aromatique est ensuite hydroxylé par la Cinnamate 4-hydroxylase (*C4H*), qui est une mono-oxygénase à cytochrome P450. Une *C4H* a été identifiée chez *Arabidopsis* et deux chez le maïs (Barrière, Riboulet et al. 2007). Deux ont été annotées chez *Brachypodium*.

Ceci est suivi par la thioestérification par Coenzyme A du groupe carboxyle, catalysée par la 4-coumarate-CoA ligase (4CL) et aboutissant à la formation de *p*-coumaroyl-CoA. Neuf 4CL ont été annotées chez *Brachypodium* mais seulement cinq ont été identifiées chez le maïs.





**Figure 30.** Voie de biosynthèse des monolignols. La voie principale est indiquée en bleu foncé, les réactions moins fréquentes en bleu clair, et les minoritaires sont sur fond blanc. (Ralph et al., 2010)

A partir du *p*-coumaroyl-CoA, trois différentes voies se distinguent : la première aboutit aux différents phénylpropanoïdes tels que les flavonoïdes, les stilbènes ou les lignanes et les lignines (Fig. 31). La seconde, après réduction par la cinnamoyl CoA-réductase (CCR), dirigera vers la production des unités H de la lignine à partir du *p*-coumaraldéhyde. La troisième, après action de la *p*-hydroxycinnamoyl-CoA/shikimate *p*-hydroxycinnamoyl transférase (HTC), ira vers les deux autres unités de la lignine à partir des acides shikimiques ou quiniques. Les deux dernières routes marquent la spécificité de la suite des réactions envers la synthèse des monolignols.

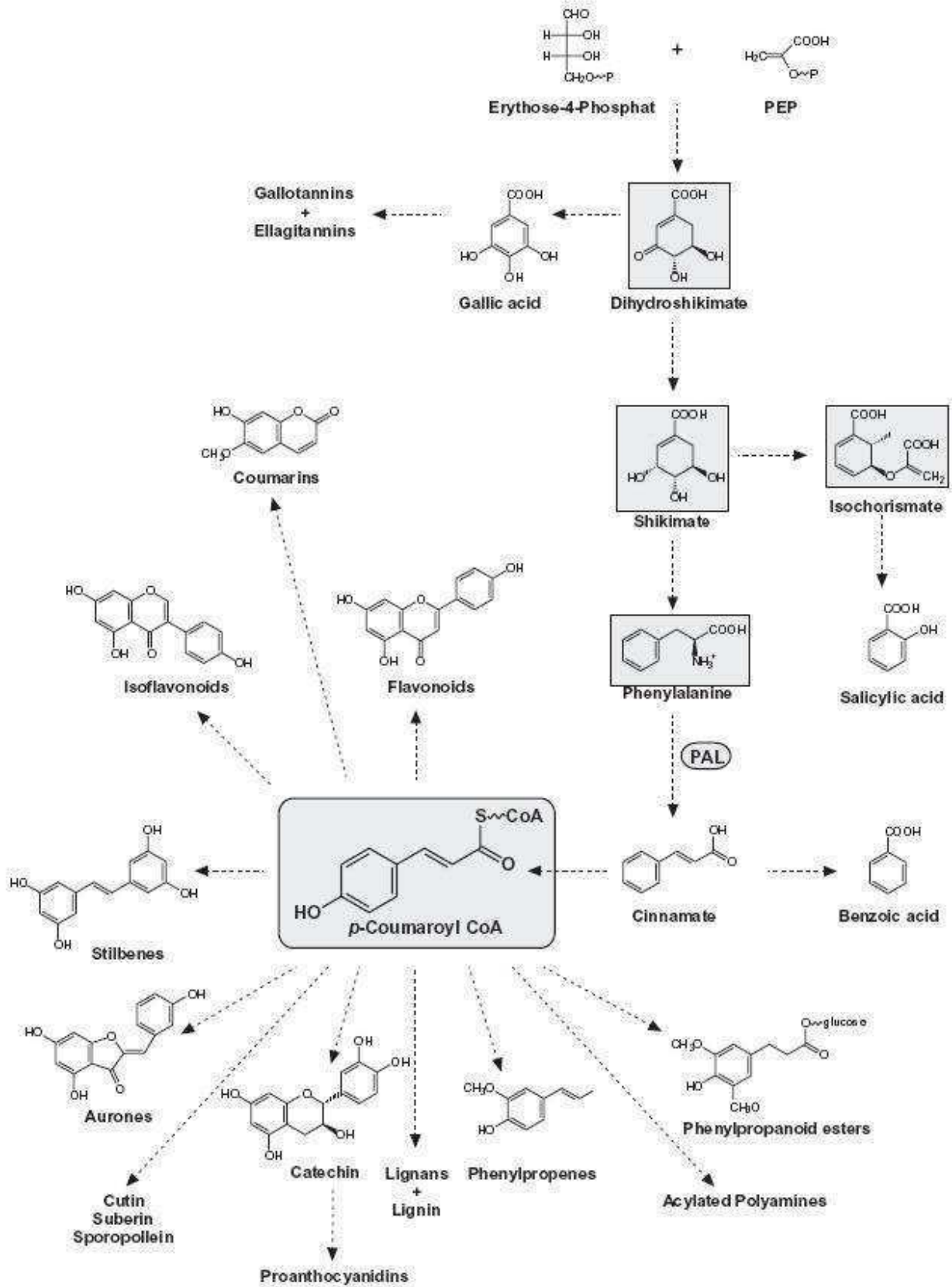
#### 4.7.3.2. Voie spécifique aux monolignols

Les acides shikimiques ou quiniques sont ensuite hydroxylés par une coumarate 3-hydroxylase (C3'H) puis convertis en cafféoyl-CoA par la HCT à nouveau (Hoffmann, Besseau et al. 2004). Les étapes suivantes sont décrites ci-après :

##### 4.7.3.2.1. Cafféoyl Coenzyme A 3-*O*-Méthyltransférase (CCoAOMT)

Les enzymes de la famille des CCoAOMT sont impliquées dans la première étape de méthylation de la voie de biosynthèse des monolignols : elles méthylent en position 3 le cycle aromatique des thioesters d'acides cinnamiques que sont le cafféoyl-CoA et le 5-hydroxyfêruloyl-CoA, produisant respectivement le fêruloyl-CoA et le sinapoyl-CoA (Fig. 32). Il a cependant été montré *in vitro* une préférence de substrat pour le cafféoyl-CoA (Parvathi, Chen et al. 2001). Lors de la réaction, le donneur de méthyle est la *S*-adénylméthionine (SAM). Les CCoAOMT impliquées dans la voie de biosynthèse des monolignols fonctionnent en dimères et nécessitent la présence d'un ion  $\text{Ca}^{2+}$  (Fig. 33) (Dixon et al, 2005).

L'implication de la CCoAOMT dans la biosynthèse des monolignols a été suggérée par les travaux de Ye (Ye, Kneusel et al. 1994) chez *Zinnia* puis chez la luzerne (Inoue, Sewalt et al.



**Figure 31 .** Voie des phénylpropanoïdes se diversifiant à partir du *p*-Coumaroyl CoA et aboutissant à la synthèse de diverses familles de composés. (Vogt et al., 2010)

1998). Les profils d'expression de ce gène sont fortement liés aux tissus lignifiés: chez les dicotylédones on retrouve des CCoAOMT spécifiquement dans tous les tissus en cours de lignification (Ye 1997). Un mutant *ccomt-1* d'*Arabidopsis thaliana* présente peu de différence dans leur quantité globale de lignine, mais on y trouve plus d'unités H et S et moins d'unités G, ce qui augmente légèrement leur rapport S/G par rapport au sauvage. Leur composition en phénols solubles est modifiée : par rapport au témoin, on y trouve seulement 50% de la quantité d'isorhamnétine 3-O-glucoside 7-O-rhamnoside et 80% de sinapoyl malate. De plus, des vaisseaux de xylème collapsés sont observables (Do, Pollet et al. 2007). Malgré son implication dans une réaction relativement précoce dans la voie de biosynthèse et impliquant la formation des unités à la fois G et S, la perte de fonction de la CCoAOMT n'a qu'un effet limité sur la quantité ou la composition des lignines. Ceci suggère qu'une ou plusieurs autres enzymes peuvent être impliquées dans cette étape de méthylation (Chen, Srinivasa Reddy et al. 2006), en particulier la COMT (Do, Pollet et al. 2007). Cependant, il a pu être observé une diminution de quantité de lignine chez plusieurs espèces sous-exprimant la *CCoAOMT*, comme chez le peuplier anti-sens (Meyermans, Morreel et al. 2000; Wei, Wang et al. 2008) ou chez le tabac (Zhong 1998).

#### 4.7.3.2.2. Cinnamoyl CoA réductases (*CCR*)

Les enzymes de la famille des Cinnamoyl CoA réductases sont impliquées dans la première réaction de la voie spécifique aux monolignols : la réduction des cinnamoyl-CoA en cinnamaldéhydes. Cette réaction nécessite le NADPH comme cofacteur. Les CCR semblent présenter une préférence de substrat pour le feruloyl-CoA (Goffner, Campbell et al. 1994; Lauvergeat, Lacomme et al. 2001; Ma and Tian 2005).

Des études d'expression chez *Lolium perenne* ont montré que chez cette monocotylédone, la CCR1, impliquée dans la lignification, atteint son maximum d'expression au stade



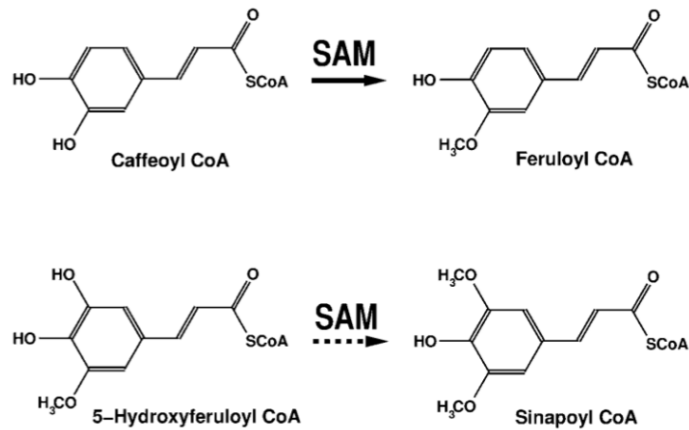
d'émergence de l'épi (Tu, Rochfort et al. 2010). Ceci est l'une des caractéristiques des gènes majeurs de biosynthèse des monolignols chez les céréales.

Il existe deux *CCR* impliquées dans la voie de biosynthèse des monolignols chez *Arabidopsis thaliana* (*CCR1* et *CCR2*). Des *CCR* ont également été identifiées chez les graminées: deux isoformes chez le maïs (Pichon, Courbou et al. 1998; Tamasloukht, Wong Quai Lam et al. 2011), une chez le ryegrass (McInnes, Lidgett et al. 2002; Larsen 2004; Tu, Rochfort et al. 2010), le riz (Lin, Yamano et al. 2006), le blé (Ma and Tian 2005; Ma 2007), le panic (Escamilla Treviño, Shen et al. 2010). La diminution ou la perte d'expression des *CCR* aboutit à une diminution du taux de lignines, des défauts de croissance et à la présence de vaisseaux de xylème collapsés (Piquemal, Lapierre et al. 1998; Chabannes, Barakate et al. 2001; Goujon, Ferret et al. 2003). Le mutant de maïs *Zmccr1* ne présente pas de défaut de développement et voit sa digestibilité améliorée grâce à une légère diminution du taux de lignines (Tamasloukht et al. 2011).

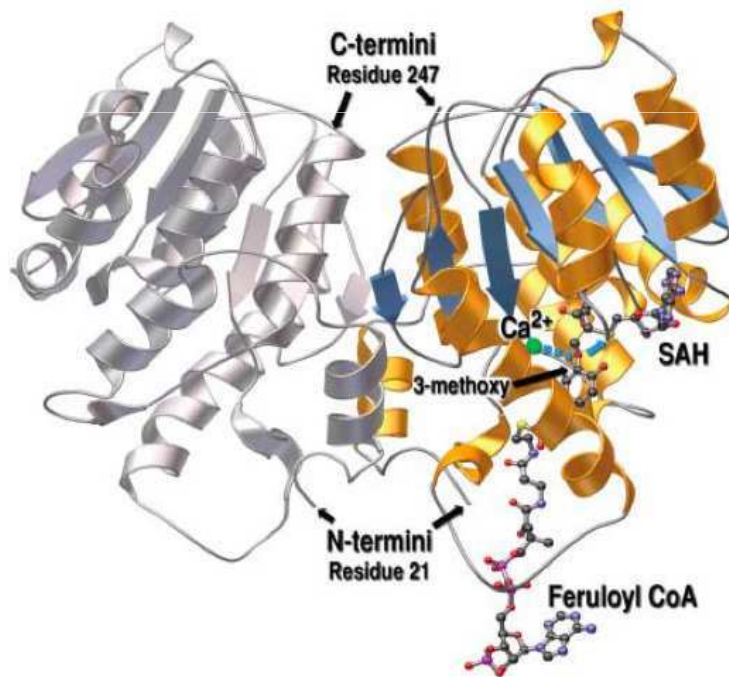
Les mutants *ccr1* d'*Arabidopsis thaliana*, en plus de leur taux de lignines n'atteignant seulement 65% à 75% de ceux du sauvage, présentent des modifications de leurs pools de phénols solubles ou flavonoïdes : ils ont trois à quatre fois moins de sinapoyl malate et accumulent du féruloyl malate (Mir Derikvand, Sierra et al. 2008). Ces résultats confirment les interconnexions entre les différentes voies du métabolisme phénoliques et les redirections de flux d'une voie vers une autre liées à des mutations.

#### 4.7.3.2.3. Férulate-5-hydroxylase (*F5H*)

La férulate-5-hydroxylase est une mono-oxygénase cytochrome P450-dépendante spécifique (famille des P450) à la biosynthèse des unités S des lignines. Elle catalyse la dernière étape d'hydroxylation du noyau aromatique, permettant la formation des précurseurs nécessaires à la formation de l'alcool synapylique, préférentiellement le coniféraldéhyde (Humphreys,



**Figure 32.** Substrats et produits de la CCoAOMT dans la voie de biosynthèse des monolignols (Dixon, 2005)



**Figure 33.** Structure dimérique 3D de CCoAOMT de Luzerne. Le monomère de droite est représenté complexé avec son substrat, son cofacteur et l'ion Ca<sup>2+</sup> nécessaires à son activité. Les structures secondaires sont indiquées par des couleurs: en bleu les feuilletts  $\beta$ , en jaune les hélices  $\alpha$ . (Dixon, 2005)

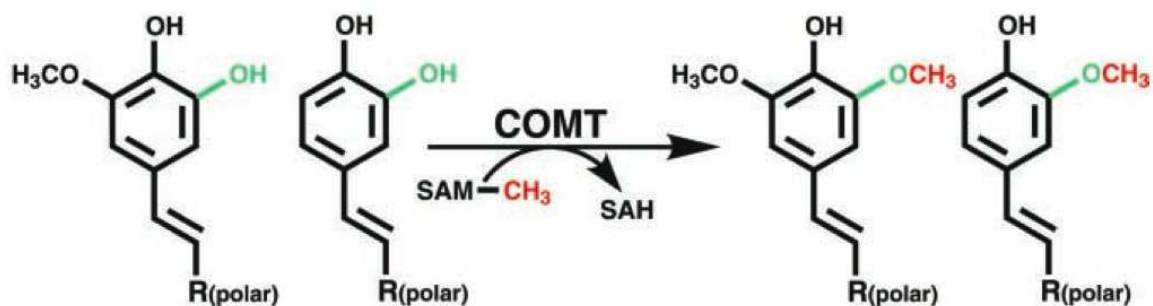
Hemm et al. 1999; Osakabe, Tsao et al. 1999) mais également l'alcool coniférylique Chen (Chen, Yasuda et al. 1999), l'acide férulique et possiblement du féruloyle-CoA, produisant respectivement le 5-hydroxyconiféraldéhyde, l'alcool 5-hydroxyconiférylique, le 5-hydroxyférulate et le 5-hydroxyféruloyle-CoA (Fig.30). Cette enzyme est spécifique aux végétaux possédant des lignines S et aucune activité F5H n'a pu être détectée chez des gymnospermes (Higuchi 1990; Fukushima and Terashima 1991), faisant de la F5H une enzyme clé des divergences évolutives entre les angiospermes et les gymnospermes. De plus, il semblerait que l'expression de la *F5H* présente une spécificité de localisation, restreinte au parenchyme interfasciculaire chez *Arabidopsis thaliana* (Meyer, Cusumano et al. 1997). La majorité des protéines de la famille des P450 possèdent un hème et nécessitent la présence du cofacteur NADPH.

Les mutants présentant une perte d'expression de la *F5H* présentent des modifications de composition de leurs lignines, contenant alors principalement des unités G et très peu d'unités S chez *Arabidopsis*, le tabac ou le peuplier (Marita, Ralph et al. 1999; Franke, McMichael et al. 2000; Stewart, Akiyama et al. 2009). Il a également été montré que la F5H d'*Arabidopsis* a une influence sur la voie des sinapates ester : le mutant *fah1* ne pouvant produire les quantités normales de 5-hydroxyférulate, peu de sinapate est synthétisé et il n'accumule pas de sinapoyl malate (Chapple, Vogt et al. 1992).

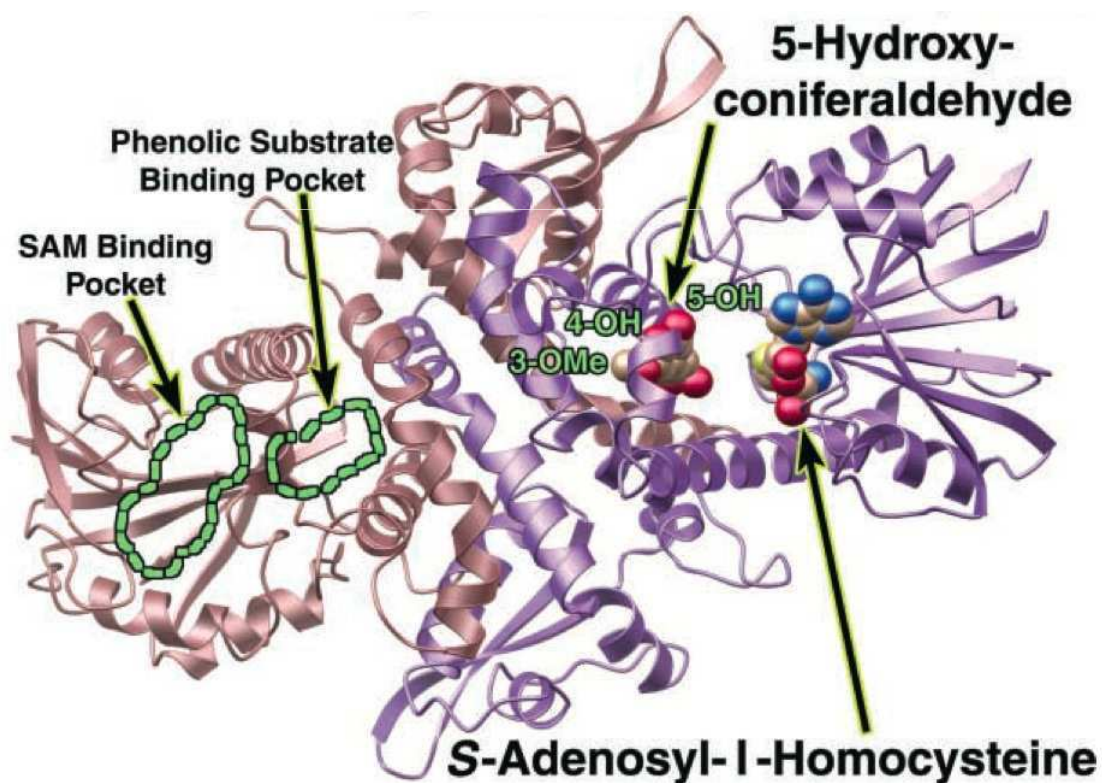
#### 4.7.3.2.4. Acide Cafféique *O*-Méthyltransférase (*COMT*)

Le rôle préférentiel des acide cafféique *O*-méthyltransférases est de catalyser la méthylation du Carbone 5 du 5-hydroxyconiféraldéhyde en sinapaldéhyde, marquant le carrefour entre la formation des deux monolignols principaux et permettant ainsi la formation des unités S de la lignine (Fig. 34). La COMT est active sous forme de dimère et nécessite la présence de SAM (la S-adénosyl méthionine) comme donneur de méthyle (Fig. 35), faisant ainsi partie de la





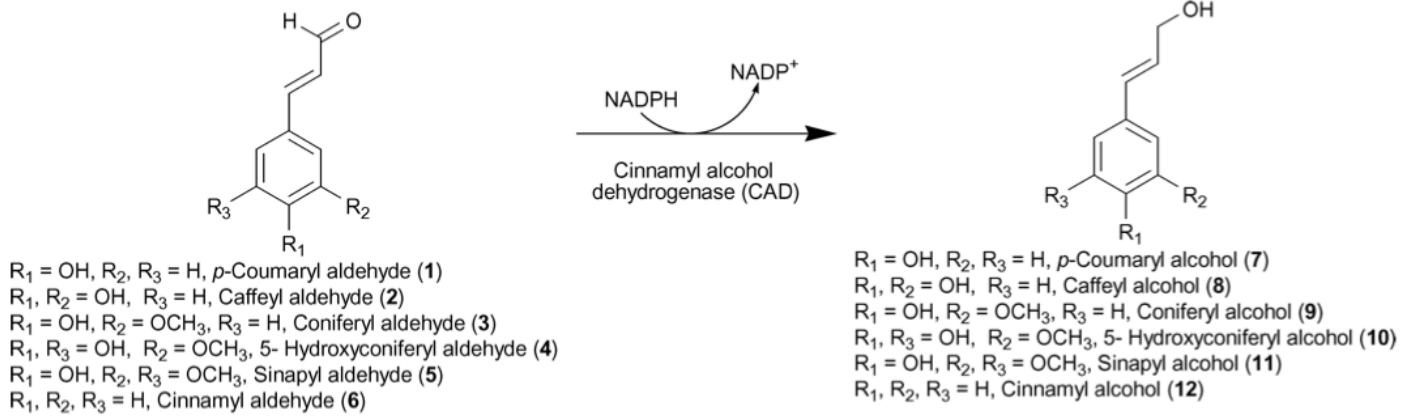
**Figure 34.** Substrats et produits de la COMT dans la voie de biosynthèse des monolignols (Zubieta et al., 2002)



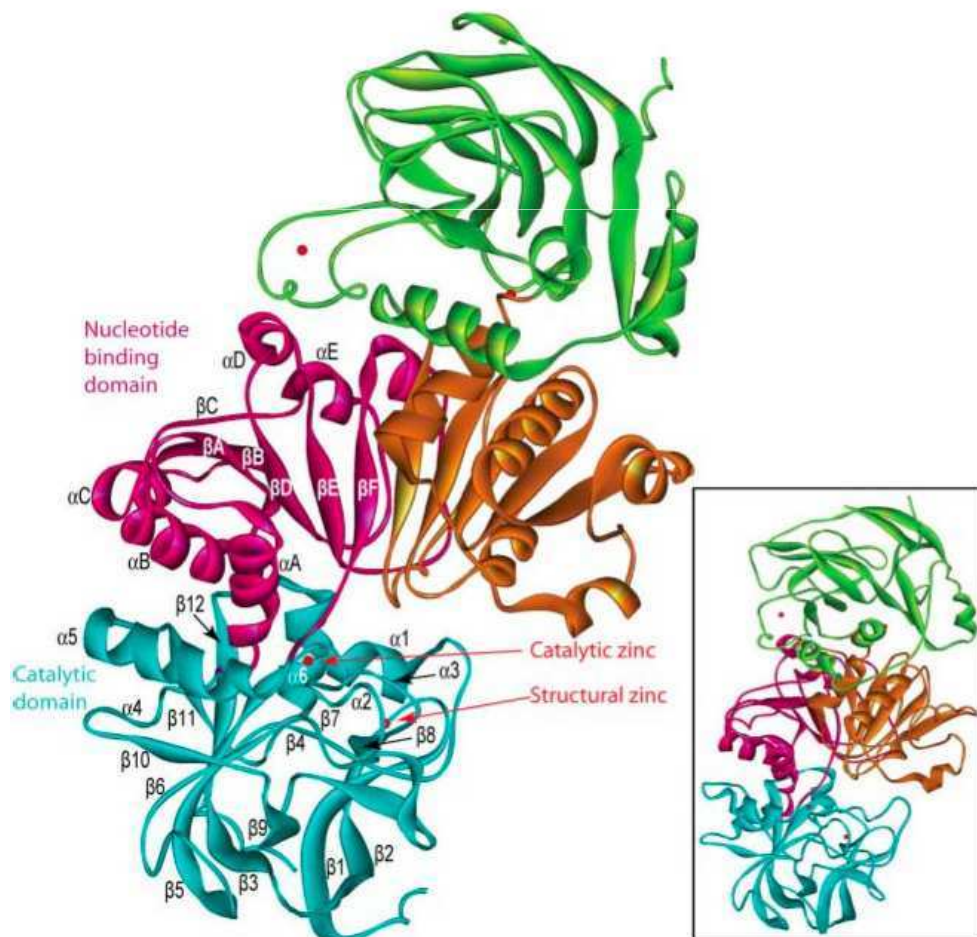
**Figure 35.** Structure dimérique 3D de COMT de luzerne. Chaque monomère est représenté d'une couleur (rose ou violette). Les poches de fixation au substrat et au cofacteur (SAM) sont indiquées en vert sur le monomère de gauche. Le cofacteur et le substrat sont montrés en 3D sur le monomère de droite. (Zubieta et al., 2002)

famille des *O*-méthyltransférases (OMT) SAM dépendantes. Deux motifs ont été identifiés comme distinguant les COMT des autres OMT (Selman-Housein and LÚpez 1999), et la cristallisation de la COMT de luzerne (Inoue, Sewalt et al. 1998; Zubieta, Kota et al. 2002) et de l'ivraie (Louie, Bowman et al. 2010) ont permis la détermination des résidus impliqués dans l'activité COMT et dans la fixation de son cofacteur.

Les plantes sous-exprimant fortement ou ayant perdu leur activité COMT présentent toutes une diminution de leur quantité globale de lignine, caractérisée par une perte quasiment totale de la présence d'unités S. L'accumulation dans les lignines d'unités d'alcool 5-hydroxyconiférylique (5-OH-G), précurseur des unités S, est également caractéristique de la perte d'expression totale ou partielle de la COMT. Il a été observé chez le maïs une augmentation de la teneur en unités G (Piquemal, Chamayou et al. 2002) cependant chez d'autres espèces telles que l'ivraie (*Lolium perenne*) ou le panic, une diminution des unités G a été constatée (Fu, Mielenz et al. 2011) tandis qu'aucune différence n'a pu être détectée chez la fétuque (Bell, Lehmann et al. 2004). L'effet sur les unités G dépend donc de l'espèce étudiée et ne peut être généralisé. La mutation dans le gène de la COMT a été liée à la coloration des lignines en rouge ou marron, particulièrement visible chez le maïs et le sorgho et au niveau de la nervure centrale des feuilles (phénotype brown midrib) (Vignols, Rigau et al. 1995; Bout and Vermerris 2003), ou une coloration rose des tiges de peuplier (Jouanin, Goujon et al. 2000). Les colorations observées pourraient être dues à l'oxydation de l'alcool 5-hydroxy-coniféraldéhyde en quinones. Mais le phénotype coloré n'a pas été observé chez toutes les espèces, comme par exemple chez *Arabidopsis* (Goujon, Sibout et al. 2003) ou le tabac (Atanassova, Favet et al. 1995). La présence d'un phénotype important chez les simples mutants suggère qu'une seule COMT serait impliquée dans la voie de biosynthèse des monolignols. Il a par ailleurs été observé chez *Arabidopsis thaliana* une modification des teneurs en flavonoïdes, particulièrement des sinapates esters (Goujon, Sibout et al. 2003) avec



**Figure 36.** Substrats et produits de la CAD dans la voie de biosynthèse des monolignols. (Noel et al., 2005)



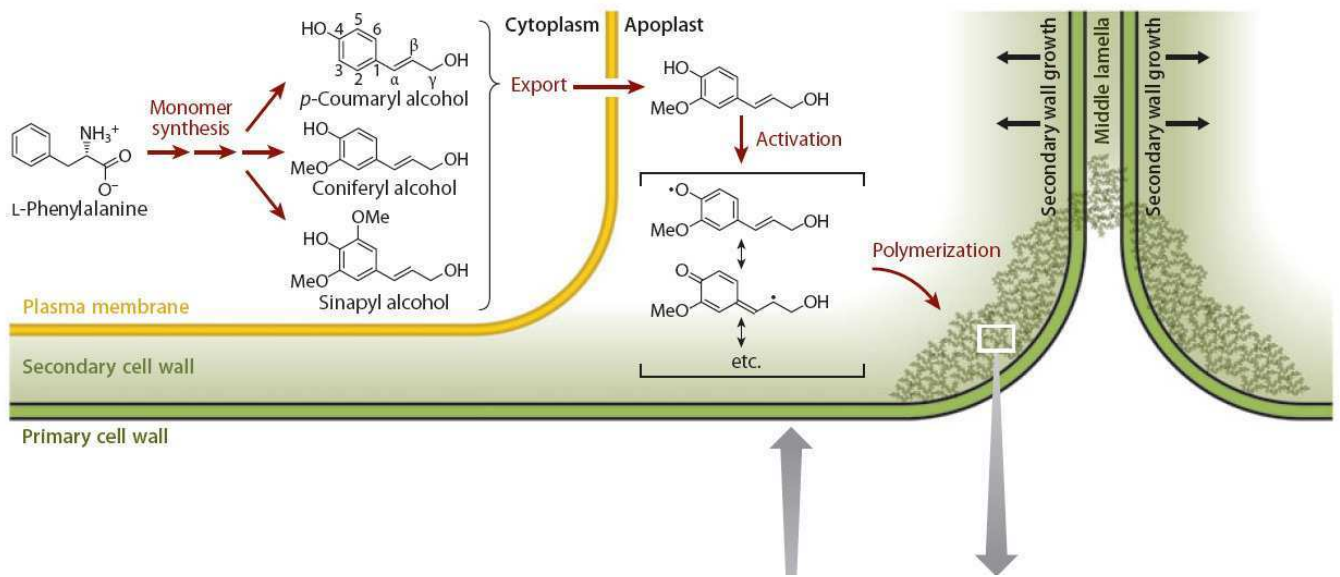
**Figure 37.** Structure dimérique 3D de CAD d'*Arabidopsis thaliana*. Chaque monomère est constitué de deux chaînes, chacune représentée dans des couleurs différentes. (Noel et al., 2005)

notamment une réduction de la quantité de sinapoyl malate, la présence de 5-OH-feruloyl malate et de 5-OH-feruloyl glucose ainsi que l'absence d'isorhamétine.

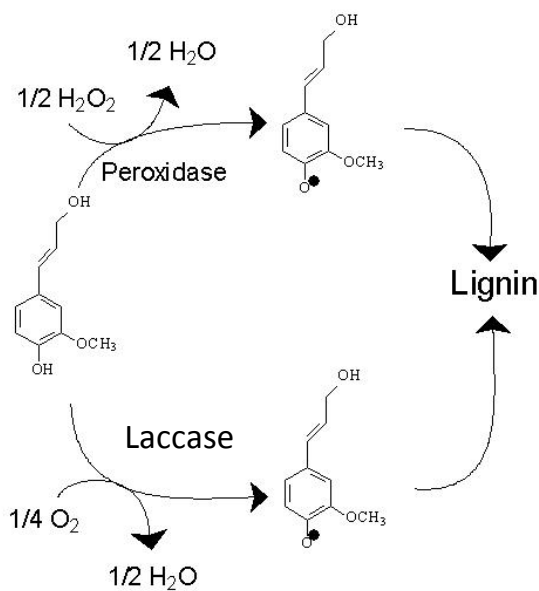
#### 4.7.3.2.5. Alcool Cinnamylique Déshydrogénase *CAD*

Les CAD impliquées dans la lignification sont des enzymes appartenant à la superfamille des MDR (medium chain dehydrogenase/reductase). Elles catalysent la dernière étape de la biosynthèse des monolignols, réduisant les cinnamaldéhydes (Fig. 36), et pour cela, ont longtemps été considérées comme un facteur limitant de la biosynthèse des lignines. L'activité CAD ayant lieu sur les trois monomères S, G et H, cette enzyme n'est pas considérée comme possédant une influence sur l'allocation du carbone dans la voie de biosynthèse (Anterola, van Rensburg et al. 1999; Anterola and Lewis 2002), celle-ci ayant lieu à d'autres niveaux (par la F5H et la COMT).

Les CAD impliquées dans la lignification fonctionnent sous forme de dimères (Fig. 37). Leur activité nécessite la présence de deux ions  $Zn^{2+}$ , l'un structural et l'autre fonctionnel, et est dépendante du co-facteur NADPH (Rossmann, Moras et al. 1974; Youn, Camacho et al. 2006). Les *CAD* forment des familles multigéniques : il en existe notamment neuf chez *Arabidopsis thaliana*, dont deux participent majoritairement à la voie de biosynthèse des lignines (Sibout, Eudes et al. 2005). Le panic possède également deux CAD majoritaires dans la voie de biosynthèse des monolignols (Saathoff, Tobias et al. 2011). Il existe par ailleurs 17 CAD chez le peuplier et 12 chez le riz (Raes, Rohde et al. 2003; Tobias and Chow 2005). L'implication des protéines de la famille des CAD dans la voie de biosynthèse des monolignols a d'abord été montrée chez l'eucalyptus par le gène EgCAD2 (Goffner, Joffroy et al. 1992) puis elles ont été caractérisées chez le zinnia, l'eucalyptus, le tabac, le peuplier, le pin, le maïs ou encore la luzerne et le sorgho, le riz, le panic ou le blé (Halpin, Knight et al. 1992; O'Malley, Porter et al. 1992; Grima-Pettenati, Feuillet et al. 1993; Somers, Nourse et al. 1995; Baucher, Chabbert et al. 1996; MacKay, O'Malley et al. 1997; Sato, Watanabe et al.



**Figure 38.** Distribution des étapes de la biosynthèse des lignines au sein de la cellule végétale. Les monolignols sont synthétisés dans le cytoplasme, puis sont exportés à la paroi où ils seront polymérisés. (Bonawitz et Chapple, 2010)



**Figure 39.** Réactions d'oxydation des précurseurs de la lignine par des laccases ou des peroxydases

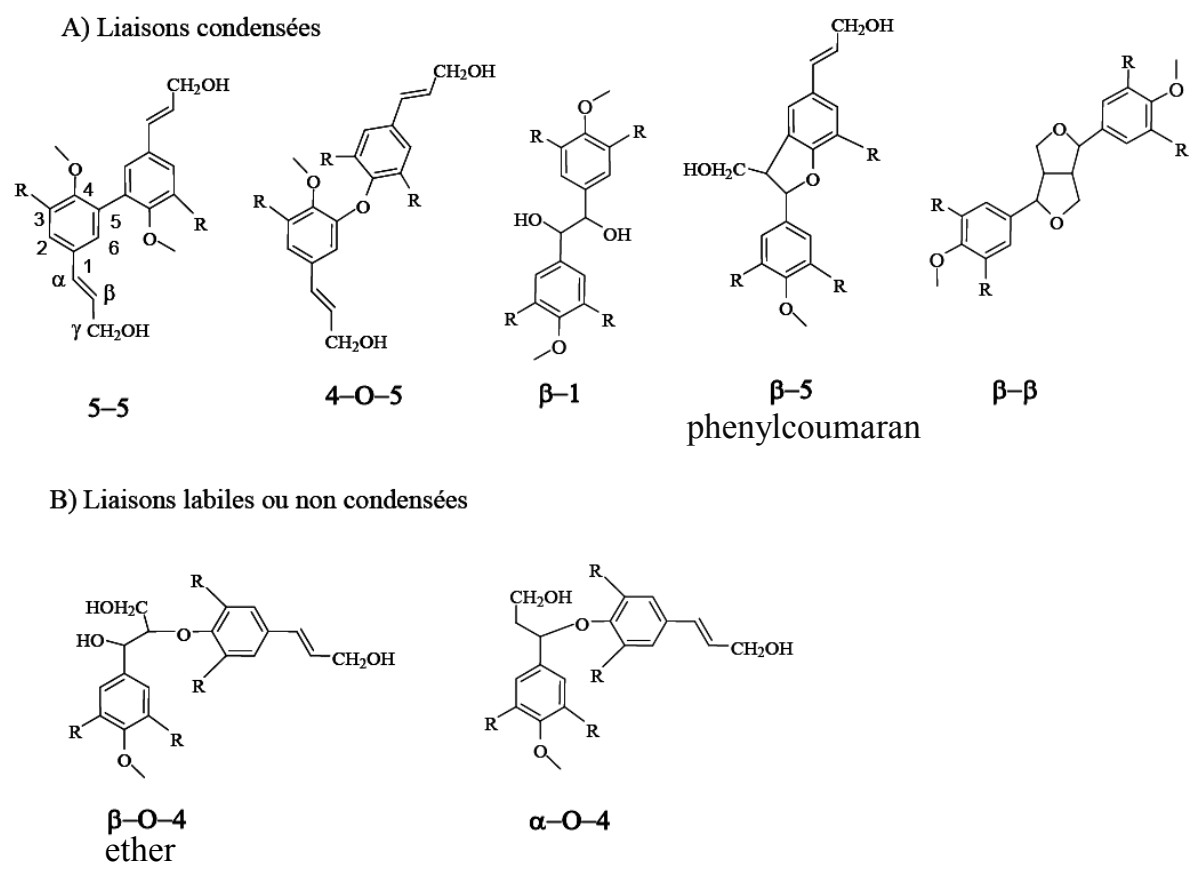
1997; Halpin, Holt et al. 1998; Baucher, Bernard-vailhé et al. 1999; Chabannes, Barakate et al. 2001; Sattler, Saathoff et al. 2009; Fu, Xiao et al. 2011).

Dans les mutants *cad*, le coniféraldéhyde peut s'inclure dans le polymère de lignine en quantités significativement plus importantes que chez le sauvage, par le même procédé de couplage que les unités G (Ralph, Lapierre et al. 2001; Kim, Ralph et al. 2002). Les lignées mutées ou possédant une ou des CAD sous-exprimées présentent une coloration rouge, brune ou orangés des tissus lignifiés, en particulier des vaisseaux du xylème. Ceci a été montré chez le peuplier (Baucher, Chabbert et al. 1996), le maïs par la lignée *bml* (Halpin, Holt et al. 1998), le sorgho par la lignée *bmr6* (Saballos, Vermerris et al. 2008), le riz par la lignée *gold hull* (Zhang, Qian et al. 2006), le panic (Fu, Xiao et al. 2011). Cette coloration est causée par des dérivés des cinnamaldéhydes accumulés dans ces lignées (Higuchi, Ito et al. 1994).

Les effets de la diminution de l'activité CAD sur la quantité et la composition des lignines dépend des espèces étudiées. Par exemple chez le riz, la lignée mutante *FCI* (flexible culm 1) présente une diminution de la quantité globale de lignine, mais pas de différences significatives de proportions de ses différents monomères (Li, Yang et al. 2009), de même que chez le tabac (Damiani, Morreel et al. 2005).

#### 4.7.3.3. Transport et Polymérisation des monolignols

Les monolignols sont des composés instables et toxiques, extrêmement réactifs. S'ils restaient sous leur forme native dans le cytoplasme où ils sont synthétisés, ils risqueraient de produire de lourds dégâts pour la cellule. Il a été envisagé l'existence d'un système de stockage et de transport des monolignols sous formes  $\beta$ -O-4 glucosylées (le glucoside d'alcool *p*-coumarylique (CAG), la coniférine, et la syringine) (Fig 38). Ceci aurait pour effet de réduire leur réactivité et d'augmenter leur solubilité, ce qui faciliterait leur export vers l'apoplasme par l'intermédiaire de sécrétions golgiennes ou d'ABC transporteurs (Samuels, Rensing et al. 2002; Ehling, Mattheus et al. 2005; Kaneda, Rensing et al. 2008). La



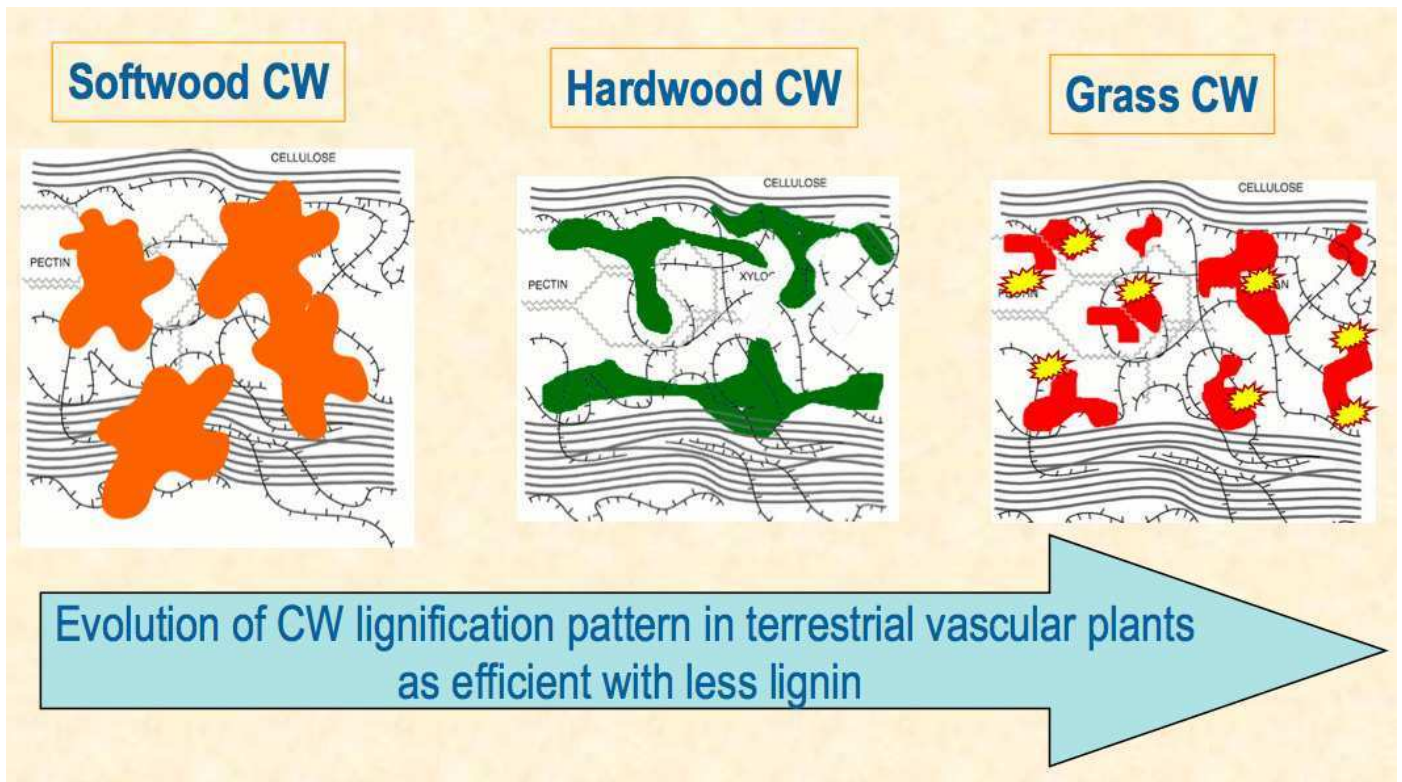
**Figure 40.** Schéma des différents liaisons possibles entre les unités des lignines.

glucosylation se ferait par des UGT (glucosyltransferases Uridine-di-Phosphate dépendantes) dont l'activité chez *Arabidopsis thaliana* a été démontrée *in vitro* (Lim, Li et al. 2001; Lim, Jackson et al. 2005) mais dont l'implication *in planta* n'a été démontrée que dans des conditions de culture génératrices de stress (Lanot, Hodge et al. 2006). La présence de monolignols glucosylés a été détectée dans des tissus en cours de lignification chez diverses espèces végétales (Dharmawardhana, Ellis et al. 1995). Une fois à la paroi, ces composés seraient alors déglucosylés par des  $\beta$ -glucosidases (Escamilla-Treviño, Chen et al. 2006) afin d'y être polymérisés (Simmons, Loquè et al. 2010).

La polymérisation des monolignols se fait par couplage oxydatif (Ralph, Lundquist et al. 2004) (Fig. 39) : les précurseurs des lignines subissent une oxydation catalysée par des laccases et/ou des peroxydases et deviennent des radicaux phénoliques. Jusque récemment, les peroxydases ont été les principales enzymes considérées comme assurant la polymérisation (Holst Christensen, Overney et al. 2001). Les peroxydases sont impliquées dans divers processus cellulaires, rendant leur rôle spécifique à la lignification difficile à déterminer mais il a été montré qu'elles sont capables de catalyser *in vitro* l'oxydation de l'alcool coniférylique (Richardson, Deighton et al. 2000; Ranocha, Chabannes et al. 2002; Demont-Caulet, Lapière et al. 2010). Cependant, des résultats récents chez *Arabidopsis thaliana* ont établi le rôle de plusieurs laccases dans la polymérisation des lignines, et présentant en particulier des spécificités des tissus et/ou de substrat (Berthet, Demont-Caulet et al. 2011). Il a été proposé que dans un premier temps les laccases pourraient permettre la formation de structures dimériques ou trimériques, qui seraient ensuite polymérisées en structures plus importantes (Dean and Eriksson 1994; Barceló 1997).

Après oxydation, les radicaux phénoliques des monomères ou oligomères seront alors spontanément couplés selon divers types de liaisons éther (C-O-C) ou carbone-carbone. Il existe cependant une autre hypothèse de polymérisation contrôlée des lignines (Gang, Costa



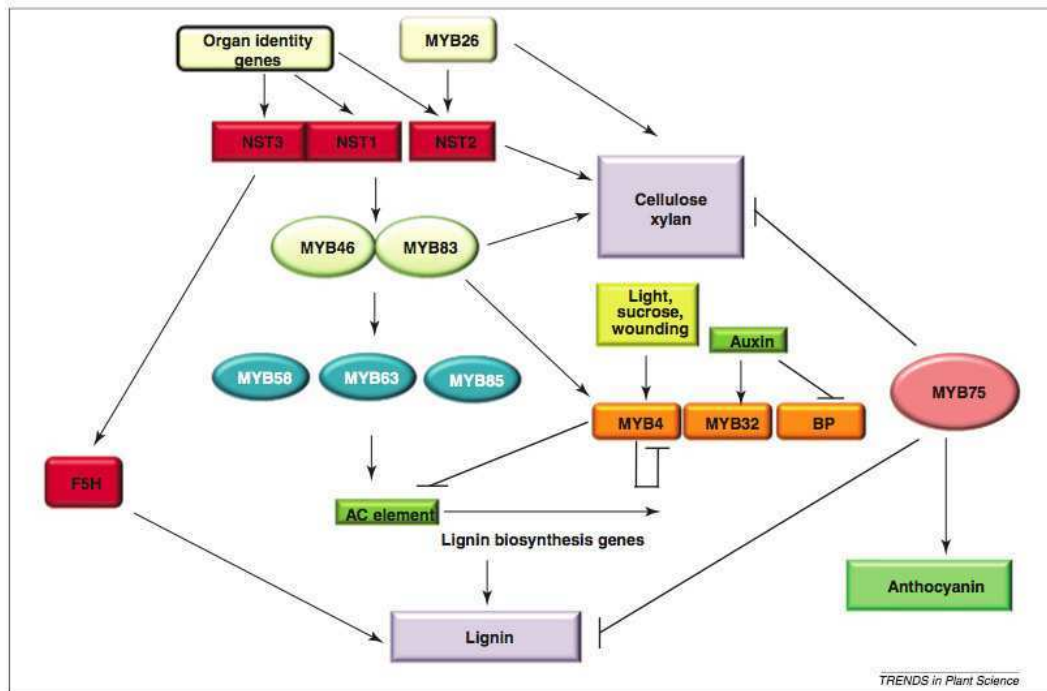


**Figure 41.** Organisation de la lignines en domaines (représentés en amas de couleurs) dans la paroi végétale. Les lignines de céréales (à droite) sont plus condensées et réparties en domaines plus petits que celles des autres espèces végétales. (Les acides hydroxycinnamiques, propres aux parois des graminées, sont représentés en jaune)

et al. 1999). Celle-ci serait dirigée par des protéines orientant les couplages radicalaires, de la même façon que dans le cas des lignanes, mais les preuves génétiques et moléculaires de leur existence restent à être établies.

Il existe deux types de liaisons interunités (Fig. 40): les liaisons labiles (principalement liaisons  $\beta$ -O-4, et  $\alpha$ -O-4) et les liaisons résistantes ou condensées (liaisons  $\beta$ -5,  $\beta$ - $\beta$ ,  $\beta$ -1, 4-O-5 et 5-5). Les monomères uniquement liés par des liaisons  $\beta$ -O-4,  $\beta$ -5 et  $\beta$ - $\beta$  formeront des structures linéaires, tandis que les liaisons 5-5 et 4-O-5 constituent des points de branchement. Les lignines riches en unités H et G sont plus condensées que celles riches en unités S. L'ensemble de ces liaisons fait que les lignines natives sont des polymères extrêmement résistants.

Les liaisons  $\beta$ -O-4 sont particulièrement importantes car ce sont celles qui sont rompues lors de dégradations industrielles (délignification pour la production de pâte à papier par voie chimique) ou analytique (analyse structurale des lignines par thioacidolyse, suivie de l'identification par spectrométrie de masse des monomères libérés). Chez les angiospermes dicotylédones, ce sont les liaisons majoritaires entre unités: dans les lignines de bois de feuillus, 60% des unités sont liées uniquement par ces liaisons (Lapierre, Pollet et al. 1995). Cependant seulement environ 20% des unités sont uniquement liés en  $\beta$ -O-4 dans les lignines de tiges de graminées. Il a été observé jusqu'à 40% d'unités liées en  $\beta$ -O-4 dans les tiges matures de *Brachypodium distachyon* (accession Bd21-3) (Catherine Lapierre, communication personnelle). Les parois de graminées sont caractérisées par la présence d'esters féruliques qui servent d'amorces de lignification (Jacquet, Pollet et al. 1995). Il est probable que ces amorces dispersées dans la paroi sont responsables de l'organisation des lignines de graminées en petits domaines dispersés et qui sont riches en groupements phénoliques libres et alcali-solubles tandis que les lignines des angiospermes dicotylédones sont organisées en plus gros domaines peu alcali-solubles et plus riches en liaisons  $\beta$ -O-4



**Figure 42.** Modèle de régulation des gènes de biosynthèse de la paroi cellulaire végétale. Les gènes de biosynthèse des lignines sont régulés par une cascade d'éléments régulateurs, à l'exception de F5H qui est directement régulé par les éléments en aval. (Zhao et Dixon, 2010)

**Tableau 8 .** Effets de modulations d'expressions de facteurs de transcription de la famille MYB sur l'expression des gènes de biosynthèse des monolignols chez différentes espèces, mesurés. (Vanholme 2010)

Transcription Factor	PAL	C4H	4CL	HCT	C3H	CCoAOMT	CCR	F5H	COMT	CAD	Method <sup>a</sup>	Reference
<i>Arabidopsis</i>												
AtMYB4	=	↓ <sup>b</sup>	↓			↑		=	=		qPCR, TEA	Jin et al.(2000)
AtMYB46	↑ <sup>b</sup>	↑	↑		↑	↑ <sup>b</sup>				↑	qPCR, TEA	Zhong et al. (2007); Ko et al. (2009)
AtMYB58	↑	↑	↑ <sup>b</sup>	↑	↑	↑	↑	↑	↑	↑	qPCR, TEA, EMSA	Zhou et al. (2009)
AtMYB63	↑	↑	↑ <sup>b</sup>	↑	↑	↑	↑	↑	↑	↑	qPCR, TEA, EMSA	Zhou et al. (2009)
BREVIPEDICELLUS	↓	↓	↓			↓ <sup>b</sup>			↓ <sup>b</sup>	↓	qPCR, EMSA	Mele et al. (2003)
<i>Tobacco</i>												
NtLIM1	↑		↑							↑	qPCR, EMSA <sup>c</sup>	Kawaoka et al. (2000)
<i>Maize</i>												
ZmMYB31		=	↓			=			↓	↑	qPCR	Fornalé et al. (2006)
ZmMYB42		↓	↓			=			↓	↓	qPCR	Fornalé et al. (2006)
<i>Poplar</i>												
PttMYB21	=	=	=			↓			=	=	qPCR	Karpinska et al. (2004)
PtrMYB3			↑			↑					qPCR, TEA	McCarthy et al. (2010)
PtrMYB20			↑			↑					qPCR, TEA	McCarthy et al. (2010)
<i>Eucalyptus</i>												
EgMYB1							↓ <sup>b</sup>			↓ <sup>b</sup>	TEA, EMSA	Legay et al. (2007)
EgMYB2	=	=	↑	↑	↑	↑	↑	↑	↑	↑ <sup>b</sup>	qPCR, EMSA	Goicoechea et al. (2005)
<i>Antirrhinum majus</i>												
AmMYB330			↓								RNA blot	Tamagnone et al. (1998)
AmMYB308	=	↓	↓ <sup>b</sup>							↓	Y1H, RNA blot	Tamagnone et al. (1998)
<i>Pinus taeda</i>												
PtMYB4	↓	=	=		↑	↑	↑		↑	↑	RNA blot, EMSA <sup>d</sup>	Patzlaff et al. (2003)
PtMYB1	↑	↑	↑		↑	↑	↑		↑↓	↑	qPCR	Bomal et al. (2008)
PtMYB8	↑	↑	↑		↑	=	↑		↓	↑	qPCR	Bomal et al. (2008)
<i>Vitis vinifera</i>												
VvWRKY2	↑	↑ <sup>b</sup>	↑		↑	↑	↑		↑	↑	qPCR, TEA	Guillaumie et al. (2010)

<sup>a</sup>qPCR refers to quantitative reverse transcription-PCR results from overexpressing lines, what does not necessarily imply direct binding on the respective promoters, except for Myb58, for which the qPCR was done in an estradiol-inducible system in the presence of the protein synthesis inhibitor cycloheximide (Zhou et al., 2009). EMSA, Electrophoretic mobility shift assay; TEA, protoplast transient expression assay; Y1H, yeast one-hybrid assay. <sup>b</sup>Direct promoter binding proven by TEA, EMSA, or Y1H. <sup>c</sup>EMSA shows binding to the Pal box. <sup>d</sup>EMSA shows binding to the AC-promoter element.

(Fig. 41). Les variations de constituants de la lignine entre les végétaux de différents groupes est également à l'origine de différences d'organisation du polymère de lignine au sein de la paroi : la présence chez les graminées de nombreux groupes phénoliques libres ainsi que des acides féruliques qui pourraient servir de points d'ancrage à la lignification suggèrent que la lignine est organisée en petits domaines dispersés (possédant de nombreux branchements dont certains terminés par des unités H qui sont peu réactives au couplage phénolique). En comparaison, les lignines de conifères sont organisées en domaines dispersés mais plus imposants, et celles des feuillus sont organisées en domaines étirés (Grabber 2005).

#### 4.7.3.4. Régulation de la biosynthèse des monolignols

La mise en place de la paroi végétale nécessite une coordination précise des gènes impliqués dans la biosynthèse, le transport et l'assemblage de ses constituants. Cependant il y a actuellement peu de connaissances sur la nature et la voie de transduction des signaux impliqués. Un certain nombre de facteurs de transcription ont été décrits comme influant sur l'expression des gènes de biosynthèse des lignines, aboutissant au modèle présenté (Fig. 42) chez *Arabidopsis thaliana* (Zhao and Dixon 2011). L'implication des facteurs de la famille des NAC (pour NAM, ATAF1, -2 et CUC2) dans la régulation des gènes de biosynthèse de la paroi secondaire, et en particulier NST1 (Nac Secondary wall Thickening promoting factor) au niveau des fibres et VND1 et VND7 (Vascular-related Nac-Domain protein) au niveau des vaisseaux, a lieu en amont de cascades d'autres facteurs de transcription (Yamaguchi and Demura 2010). Ces derniers facteurs comprennent notamment des membres des familles LIM (LIN11, ISL-1 et MEC-3) et MYB, en particulier les MYB58, MYB63 et MYB85, qui sont exprimés dans le xylème où ils se fixent sur les séquences AC des promoteurs des gènes de biosynthèse des monolignols et régulent leur expression. MYB85 est associé à la régulation des gènes de biosynthèse des monolignols, mis à part *F5H* dont l'expression est directement



régulée par les facteurs en amont (Zhao, Wang et al. 2010). Chez le maïs, les facteurs ZmMYB31 et ZmMYB42 ont des effets directs sur l'expression des gènes *C4H*, *4CL*, *CCoAOMT*, *COMT* et *CAD* (Fornalé, Sonbol et al. 2006) (Tableau 8). Cependant, ces systèmes de régulation semblent être bien plus complexes et reliés à d'autres fonctions biologiques que la lignification (Zhao and Dixon 2011).



## Objectifs du travail de thèse

Les laboratoires accueillant ce projet de thèse sont impliqués dans le projet européen RENEWALL, visant à déterminer les facteurs influençant la biosynthèse des parois cellulaires végétales dont les modifications permettraient l'amélioration de la biomasse lignocellulosique, et notamment de son potentiel de saccharification, dans le but de l'utiliser en tant que matière première renouvelable. L'application principale des parois ainsi modifiées serait la production de bioéthanol de seconde génération. Cette thèse s'inscrit dans le Work Package 2 qui a pour tâche de développer la compréhension de la biosynthèse des lignines et de ses modifications aboutissant à l'optimisation de la saccharification.

Ce travail de thèse a consisté en l'identification de gènes impliqués dans la voie de biosynthèse des monolignols chez *Brachypodium distachyon*, en se basant sur une stratégie de gènes candidats. Quatre gènes en particulier ont été ciblés : ceux codant pour la Cafféoyl Coenzyme A 3-*O*-Méthyltransférase, pour la Férulate-5-Hydroxylase, pour l'Acide Cafféique *O*-Méthyltransférase et pour l'Alcool Cinnamylique Déshydrogénase. Les trois derniers gènes sont les plus en aval de la voie de biosynthèse des monolignols et présentent l'intérêt que leur perte de fonction entraîne des modifications importantes de quantité et/ou de composition en lignines, mais sans impact majeur sur le développement et la biomasse des lignées considérées.

Les gènes candidats ont été déterminés *in silico* sur la séquence génomique complète de l'accession Bd21, disponible depuis 2008. Leur fonction a pu être vérifiée par complémentarité fonctionnelle de mutants d'*Arabidopsis thaliana* leur correspondant. Ceci a été conduit conjointement avec la recherche de mutants dans la collection de lignées mutagénisées (à l'azoture de sodium et à l'EMS) de *Brachypodium*, mise en place à l'INRA





de Versailles en collaboration avec l'URGV d'Evry. L'exploitation de cette population de mutants a été menée selon deux stratégies :

- La première, par génétique classique, a consisté en l'identification de lignées présentant des phénotypes potentiellement liés à des défauts de lignification, en particulier la couleur des tiges, suivie de l'identification d'un gène impliqué dans ce trait phénotypique.
- La seconde, par génétique inverse, s'est basée sur la technique de TILLING (Targeting Induced Local Lesions in Genomes) pour identifier des lignées mutantes dans les gènes candidats et analyser leurs effets sur la composition en lignines des parois.

La concordance des résultats des analyses bioinformatiques et fonctionnelles a permis de valider trois gènes candidats.



## Chapitre II :

### PROJET DE PUBLICATION

Disrupting the cinnamyl alcohol dehydrogenase 1 (*BdCAD1*)  
gene leads to altered lignification and improved  
saccharification in *Brachypodium distachyon*

## CHAPITRE II : Disrupting the cinnamyl alcohol dehydrogenase 1 (*BdCAD1*) gene leads to altered lignification and improved saccharification in *Brachypodium distachyon*

Ce chapitre présente les travaux effectués sur le gène de cinnamyl alcool déshydrogénase *BdCAD1*, sous forme d'une publication en cours de soumission au journal *New Phytologist*. Il décrit les résultats de la campagne de phénotypage effectuée sur la collection de mutants de *Brachypodium* mise en place par mutagenèse chimique à l'azoture de sodium et à l'EMS à l'INRA de Versailles en collaboration avec le génopôle d'Evry.

## Disrupting the cinnamyl alcohol dehydrogenase 1 (*BdCAD1*) gene leads to altered lignification and improved saccharification in *Brachypodium distachyon*

Madeleine Bouvier d'Yvoire<sup>1</sup>, Oumaya Bouchabke-Coussa<sup>1</sup>, Leonardo Gomez<sup>2</sup>, Olivier Darracq<sup>1</sup>, Wannes Voorend<sup>3</sup>, Laurent Cézard<sup>1</sup>, Philippe Lebris<sup>1</sup>, Sylvain Legay<sup>1</sup>, Catherine Lapierre<sup>1</sup>, Lise Jouanin<sup>1</sup>, Richard Sibout<sup>1</sup>

<sup>1</sup>Institut Jean-Pierre Bourgin, UMR1318 INRA-AgroParisTech, INRA Centre de Versailles-Grignon, Route de St- Cyr (RD10), 78026 Versailles France. <sup>2</sup> Centre for Novel Agricultural Products, University of York, York YO10 5 YW, UK. <sup>3</sup>Institute for Agricultural and Fisheries Research, Plant Sciences Unit - Growth and development, Caritasstraat 21, 9090 Melle, Belgium.

Author for correspondance :

Richard Sibout

Tel : +33 1 30 83 37 51

Email : richard.sibout@versailles.inra.fr



## Summary

- Second generation biofuels using fermentable sugars from plant cell walls, mainly cellulose, could provide an answer to the augmentation of global energy needs with a limited impact on food resources. However lignins constitute an obstacle to most industrial processes targeting cell wall polysaccharides, and plants with altered lignification often display enhanced saccharification yields.
- In order to identify cell wall compositions more susceptible to saccharification in *Brachypodium distachyon*, the model for grass species, the sodium azide and EMS mutagenized lines from the INRA-Versailles collection were screened based on the color of their stems, then characterized.
- The mutants identified, hereafter named “*brown stems*” (*bs*), were analyzed and several lines showed significantly improved saccharification. Among them, mutants in a cinnamyl alcohol dehydrogenase gene *BdCAD1* had the highest saccharification yields. These lines display a significant decrease in CAD activity and lignin content. Their lignins are also enriched in free phenolic groups and sinapaldehyde units and are linked by a high proportion of resistant bonds. Expression analysis and functional complementation of *Arabidopsis thaliana* and *Brachypodium distachyon cad* mutants confirmed the function of *BdCAD1*.
- This work demonstrated that *BdCAD1* is the main gene from the cinnamyl alcohol dehydrogenase family involved in the monolignol biosynthetic pathway in *Brachypodium distachyon*.

## Keywords

*Brachypodium distachyon*, cell wall, lignin, lignification, cinnamyl alcohol dehydrogenase (CAD), saccharification





## Introduction

Bioenergies derived from lignocellulosic cell walls are considered as promising sustainable alternatives to fossil fuels. These second generation biofuels, using fermentable sugars from cell walls of low input plant biomass, could provide a solution to the evolution of energy needs without drawing on food resources (Yuan, Tiller et al. 2008). However, it has been shown that lignin is one of the main factors of cell wall recalcitrance to saccharification and this phenolic polymer needs to be removed. It has been reported that modifications of lignin content, composition and structures impact digestibility and fermentable sugar yields by enzymatic degradation (Chen and Dixon 2007). Lignin polymers are major components of secondary cell walls, ranging from 7 to 15% of grass cell wall (Kuc and Nelson 1964; Ishii 1997). They derive from monolignols: *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, the polymerization of which generate H, G and S lignins respectively. The composition and quantity of lignins depend on intrinsic and environmental factors. These polymers hold a great importance in terrestrial plants as lignin confers hydrophobicity and rigidity to plants cell walls, enabling upright growth and water conduction, as well a response to stress and pathogens. As a result, overly modified lignin composition or quantity could interfere with plant development, which leads to sometimes significant biomass reduction (Mir Derikvand et al. 2008).

In order to identify cell wall compositions less refractory to saccharification, especially through changes in lignin content, but without significant decrease in lignocellulosic biomass, the *Brachypodium* mutant collection of Versailles (<http://urgv.evry.inra.fr/UTILLdb> + Dalmais et al., unpublished) was screened for phenotypes potentially linked to lignin modifications. Reddish-brown pigmentation of grass stems has been empirically linked to lignification for decades in maize (Kuc and Nelson 1964; Cherney, Mertens et al. 1990; Cherney, Cherney et al. 1991; Halpin, Holt et al. 1998). This specific phenotype (*bm* for brown midrib) has also been



observed in rice (Zhang, Qian et al. 2006) sorgho (Dien, Sarath et al. 2009; Saballos, Ejeta et al. 2009) millet (Hartley, Morrison III et al. 1992) or barley (Halpin *et al.*, unpublished). In many *bm* lines, digestibility of the cell wall is improved (Cymbaluk, Gordon et al. 1973; Li, Weng et al. 2008; Vermerris, Sherman et al. 2010). Among the lines displaying the *bm* phenotype, several are related to cinnamyl alcohol dehydrogenase (CAD) deficiencies.

Enzymes from the CAD family belong to the medium chain dehydrogenase/reductase (MDR) superfamily. They catalyse the last step of monolignol biosynthesis, from cinnamaldehydes to cinnamyl alcohols (Boerjan, Ralph et al. 2003). Mutants with a loss of CAD activity are known to have altered lignin structures, such as a decrease of the S/G monomer ratio and the incorporation of specific precursors (Kim, Ralph et al. 2002), as well as a decrease in overall lignin content in the *bm1* maize line (Halpin, Holt et al. 1998), *Arabidopsis* (Sibout, Eudes et al. 2003) and the *bmr6* sorgho line (Saballos, Ejeta et al. 2009). The *CAD* family in *Brachypodium* consists of 7 members (Vogel, Garvin et al. 2010) and according to data in other species, one or two of them should be mainly involved in lignification. The CAD proteins involved in lignification have been shown in *Arabidopsis thaliana* to function in dimers, need NADP as cofactor and are  $Zn^{2+}$  dependent (Youn, Camacho et al. 2006).

The assessment of data in *Brachypodium distachyon*, the recently established model for temperate grasses (Draper 2001; Opanowicz, Vain et al. 2008; Vogel and Bragg 2009), should help understand how lignins can be modified, either in quantity or quality, to make the isolation and utilization of cellulose from dedicated grass easier without affecting their biomass.



## **Materials and Methods**

### **Brachypodium phenotyping**

*Brachypodium distachyon* plants from the Bd21.3 accession were used in this work. Seeds were stratified in water in the dark at 4°C for 3 days then passed for 10 days in greenhouse conditions : 16/8 light/dark time temperature 25°C, night temperature 15°C. The plants were vernalized for 3 weeks under 8h light and 10°C conditions, then kept for 2 months in the greenhouse conditions until maturation and eventually dried for 1 month under the same conditions but without watering.

Phenotyping based on the color of the plants was performed visually and confirmed by analysis of data acquired on grinded dry matter with the digital color meter included in Mac OSX. ACP was done using micosoft excel and clustering based on color was done using TMeV v4.3 with default parameters.

### **Brachypodium plant material and growth conditions for analysis**

Mutants and wild type Bd21.3 were grown in growth chambers for 4 weeks under short day (8h), 21°C day, 18°C night and 70% humidity, then transferred to long day (16h) with the same temperature and humidity conditions.

The mutant lines were named INRA-V followed by a number identifying each line.

### **Saccharification assays**

Assays were conducted following the method described by Gomez et al. (2010) in which the samples were prepared using the grinding and loading robot (Labman Automation Ltd, <http://www.labman.co.uk>) then formatted in 96-well plates to contain 4 mg of each sample either as four or eight replicates. The Tecan (<http://www.tecan.com>) automated liquid handling station was used to perform the following steps: a 0.5N NaOH pre-treatment followed by 6



rincing stages with 500 µl sodium acetate buffer on the grinded material, then hydrolysis by incubation of the samples at 50°C for 8 hours in the presence of enzyme cocktail (4:1 ratio of Celluclast and Novozyme 188) in proportions of 6.3 filter paper unit per gram of material. Automated sugar determination was then carried out as previously described (Gomez, Whitehead et al. 2010).

### **Lignin analysis**

Dry plant material was sorted as to only keep the stems, grinded, and free phenolic compounds were extracted using a Soxhlet apparatus with toluene:ethanol (2:1v/v), ethanol then water sequentially. Klason procedure as described by (Dence 1992) was performed on the extract-free samples for lignin content determination. Lignin composition was identified by thioacidolysis procedure as described by (Lapierre, Monties et al. 1985) on previously extracted samples or directly on the grinded material, followed by GC-MS analysis.

### **Phylogenetic analysis**

Sequences of known or putative CAD proteins were aligned on the web server Multalin (<http://multalin.toulouse.inra.fr/multalin/>) after adjusting the output parameters to our needs, then phylogenetic analysis was performed on the same sequences on the MABL web server phylogeny.fr (<http://www.phylogeny.fr/>) using the default settings of the “one click” mode.

### **3D Structure predictions**

3D structure of BdCAD1 was predicted from the structure of cristalized AtCAD5 (PDB sum identification: 2cf5) by the web server of the molecular modelization program Geno3D (<http://geno3d-pbil.ibcp.fr/>). Coordinate files generated were visualized and edited using MacPymol and Pymol (<http://www.pymol.org/>).





## **Crossing**

Crossings were performed as described by Garvin (<http://www.ars.usda.gov/>)

## **RT-PCR**

Plant material was harvested in liquid nitrogen 4 hours after daylight switch in the growth chambers to match the maximum expression of lignin genes and grinded. RNA was extracted using the Qiagen RNeasy minikit with the additional RNase step, dosed with a Nanodrop (Thermo Scientific®), and 1µg was retro-transcribed with the Invitrogen SSRTII kit using the manufacturer's recommendation. PCR was performed using Taq from Invitrogen with a denaturation temperature of 95°C, or Finnzyme Phusion with a denaturation temperature of 98°C, and in both cases for 35 cycles with 60°C hybridation and 20 seconds elongation.

## **Enzymatic activity assays**

Stems (nodes + internodes) were harvested in liquid nitrogen 4 hours after daylight switch and grinded in liquid nitrogen. Total protein content was extracted (Tris-HCl 0,1 M pH7,5; PEG 6000 0.5%; PVP 0.4%; β-mercaptoethanol 15mM 0,1%) and quantified by Bradford. CAD activity was determined on 3 biological repeats with 3 technical repeats for each sample by colorimetric kinetics at 405nm every 40 seconds for 1h 30min, using 20µg protein extract, 50µL coniferylic alcohol 10mM, 25µL NADP 10mM and Tris-H Cl 0.1mM pH8.8 and the corresponding controls: without NADP to substract non CAD dehydrogenase activity or without protein extract to substract natural oxydation.

## **Cloning and preparation of transformation vectors**

*BdCAD1* was amplified by PCR from Bd21.3 cDNA using primers flanking the coding sequence (ATGGGCAGCATCGCCTCCGA and TCAGGCGGCATCCTCGATCTTG) and adding Gateway ATTB boxes. Gel purified products with NucleoSpin<sup>®</sup> Extract II kit (from



Macherey-Nagel) were ligated into the pDONR207 plasmid by BP Gateway<sup>®</sup> (Invitrogen) reaction then electroporated into *E. coli* top10 bacteria. Transformed clones were selected on gentamycin containing medium, their plasmids were extracted with the Nucleobond<sup>®</sup> kit (Macherey-Nagel) and sequenced. These plasmids then underwent a LR Gateway<sup>®</sup> (Invitrogen) reaction for the *BdCAD1* coding sequence to be ligated into the pGWB-2 plasmid, under control of the CaMV 35S promoter. These plasmids were electroporated into *E. coli* top10 and selected on LB growth medium containing kanamycin. Plasmids of the selected clones were extracted, checked by PCR using a left primer in the 35S promoter and a right primer in the *BdCAD1* sequence (primer sequences table supplementary data), and electroporated in *Agrobacterium tumefaciens* (C58PMP90).

### **Arabidopsis transformation**

*Arabidopsis thaliana cadc-cadd* mutants from the Wassilewskija accession (Sibout, Eudes et al. 2005) were grown in growth chambers and transformed by flower infiltration (as described by (Bechtold and Pelletier 1998). Transgenic plants were selected on MS growth medium containing hygromycin at a concentration of 50mg/L.

### **Histology**

Stems (lowest internode) were harvested and fixed in 70% ethanol for a minimum of one week. Samples were rehydrated in water overnight and embedded in 8% agarose in order to make 100µm stem cross sections using a vibratome (Leica VT1000S ). Weisner staining was achieved by incubating cross sections in phloroglucinol-HCl (VWR Prolabo) for 10 minutes. Sections were mounted in water and immediately observed under a Leica DMRB microscope equipped with a Progres C10plus camera. Maüle staining was performed by first incubating sections in KMnO<sub>4</sub> (1%). After 15 min, sections were washed in water, acidified with HCl

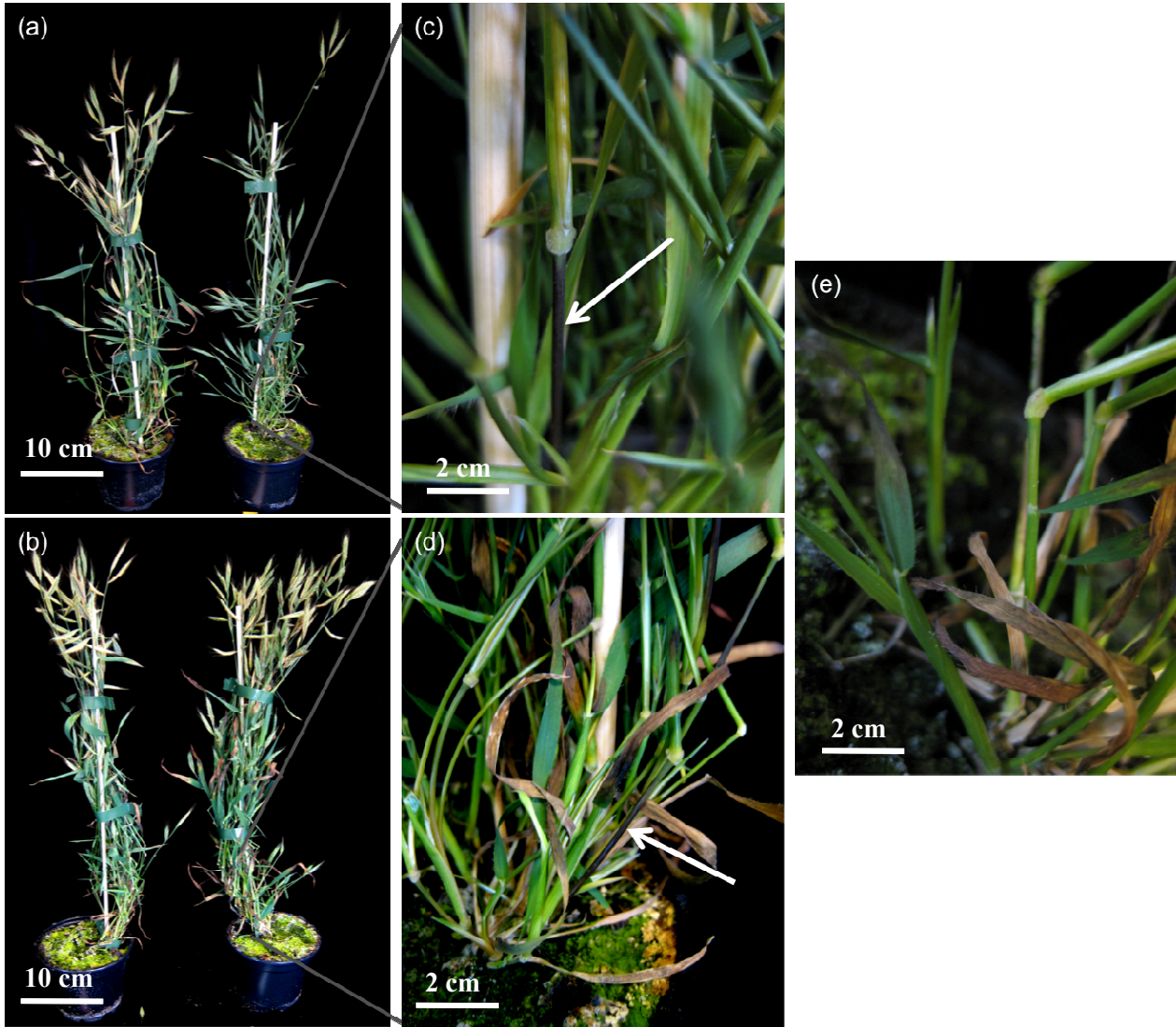


Figure 1. Visual phenotyping of *brown stem* mutants: the example of two *bs* lines, (a) and (b) right plants, were selected on their color and on their height or apparent biomass being similar to that of the wild type lines (a) and (b) left plants. The brown stem phenotype is the most visible (arrows) on lower internodes of the mutant lines, (c) and (d), compared to the wild type Bd21-3 (e).

(30%) for 1 min, washed again, and finally incubated in NaHCO<sub>3</sub> (5%) for 10 minutes before being observed under the same microscope as above.

### **Brachypodium transformation**

*BdCAD1* was cloned in *Brachypodium* and cloned in pIPKb2 vector (Himmelbach, Zierold et al. 2007). Transformation method was adapted from (Vogel, Garvin et al. 2006) and (Cherney, Cherney et al. 1991; Alves, Worland et al. 2009).

Immature embryos from the 4179 mutant were harvested at the right stage (still very small, translucent and 0.2 to 0.6mm in length) and put on callus induction medium (Linsmaier & Skoog, sucrose 30g/L supplemented with 2.5 mg/L 2,4D and 0.6 mg/L CuSO<sub>4</sub>) for 3 weeks to produce embryogenic callus. Calli were then subcultured during 2+1 more weeks on the same medium for multiplication. After multiplication, calli were inoculated with *Agrobacterium tumefaciens* strain AGL1 containing the pIPKb2 ::BdCAD1 binary vector. *Agrobacterium* was grown on LB solid medium supplemented with streptomycin 25mg/L, spectinomycin 50mg/L, rifampicin 25mg/L and carbenicillin 50mg/L and then suspended in liquid LS medium containing 60mg/L Acetosyringone (Sigma # D134406) and 10% Synperonic PE/F68 (Sigma #81112) until a DO<sub>600</sub>=0,6. Calli were inoculated for 5 min and then co-cultured on dry filter paper at 24°C during 3 days. They were then transferred on selection medium during 3 weeks (LS, sucrose 30g/L supplemented with 2.5 mg/L 2,4D, 40mg/L Hygromycin B, 250mg/L Timentin and 0.6 mg/L CuSO<sub>4</sub>) and calli transferred to regeneration medium (LS, maltose 30mg/L, supplemented with 0.2 mg/L kinetin, 40mg/L Hygromycin B and 250mg/L Timentin). After 2-3 weeks, small shoots, with or without small roots were transferred in tubes on rooting medium (half-strength Murashige & Skoog, Sucrose 10g/L) for 2 weeks. Fully rooted plantlets were then transferred to the greenhouse after 2 weeks of vernalisation at 4°C in compost pots for seed production.

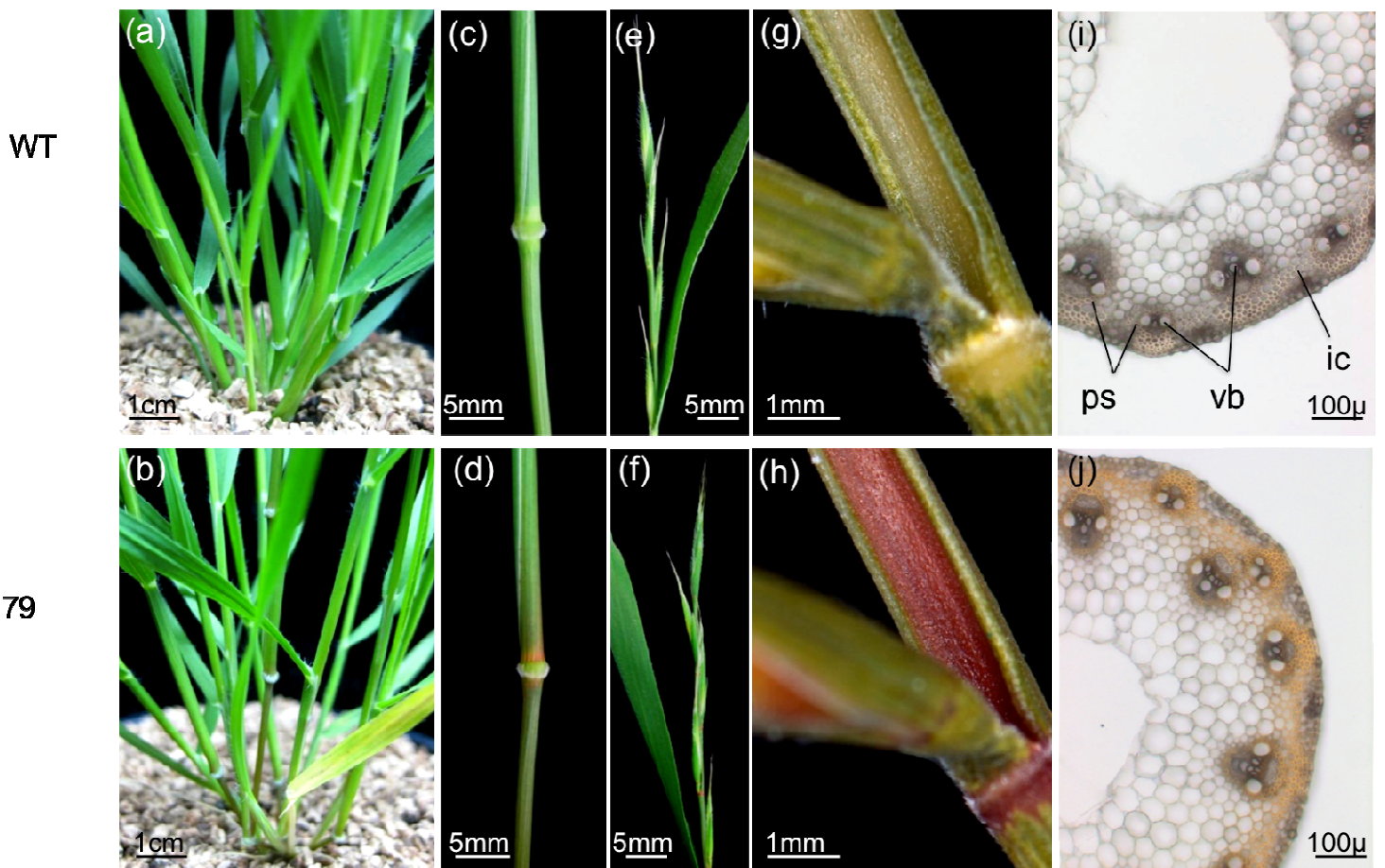


Figure 2. Visual phenotyping of a particular *brown stem* mutant: line 4179. The brown stem phenotype is the most visible at the flowering stage (9 weeks). It is more present on lower internodes when comparing the wild type Bd21-3 (a) to the 4179 line (b) and a red/brown coloration is particularly noticeable around the nodes (c) and (d) and at the base of the spikelets (e) and (f). This coloration is markedly intense on the rachilla (g) and (h). Unstained sections of the top internode (i) and (j) show this coloration is located in the lignified tissues (vb, vascular bundles; ps, perivascular sclerenchyma; ic, interfascicular collenchyma)

## Results

### ***brown stem* lines show improved saccharification**

To identify lines with altered lignins, the phenotyping campaign conducted on the Versailles Brachypodium mutants (Dalmais *et al.* unpublished) was partly focused on color phenotypes of stems and leaves, in particular any shade of red, brown or orange. As the main purpose was to point out traits leading to an improvement of saccharification yields, only lines showing no or very little decrease in biomass were listed (Fig. 1a, b). This led to the isolation of twenty one lines that were thus candidates for improved saccharification. The coloration being mostly visible on the stems, this phenotype is now referred as *brown stem* or *bs* (Fig. 1c, d, e). Most lines kept their color on exposed areas even after maturation. However in several lines the pigmentation of the stem was the deepest around the flowering stage, then faded during maturation, as observed on the *bm* maize mutants (Cherney, Cherney et al. 1991). Among them, the lines 4179 and 7591 were identified and described in this work. They displayed a brown/dark red coloration, mostly on their lower internodes (Fig.2 a, b) and around their nodes (Fig. 2 c, d), but that can also be observed all along the internodes following the collenchyma positioned behind the vascular bundles. The basis of their spikelets (Fig. 2e, f) and the rachilla facing the florets also displayed an intense red coloration (Fig. 2g, h). This coloration was the most visible at the time and following flowering, before complete maturation.

Automated saccharification assays were conducted in order to assess the modifications in sugar release yields in the *bs* mutant lines (Fig. 3).

Only one line displayed a slight decrease in saccharification yield at 94.5% of that of the WT. The other lines had equal or superior yields, with maximum values at 143.7% and 152.8% of the WT for the 7591 and 4179 lines respectively.

Klason and thioacidolysis lignin analysis on the *bs* lines with the highest increase in saccharification yields (Table 1) showed a significant decrease in total lignin contents. The two



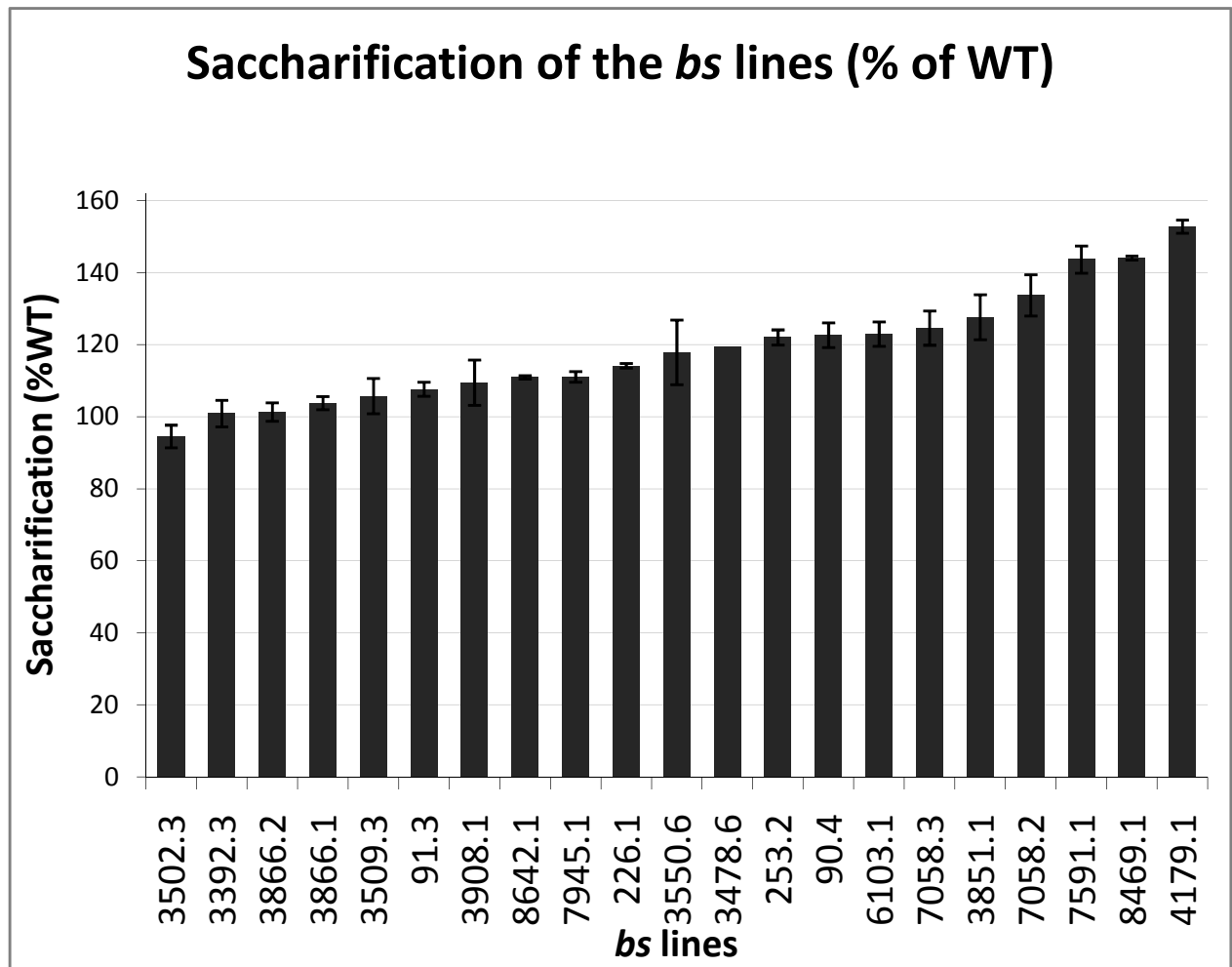


Figure 3. Saccharification assays on *bs* lines, expressed in % of the saccharification yields (in nmol/gmaterial.hour) of the corresponding controls. Lines are ordered according to the efficiency of the saccharification kinetics. Only the line 3502.3 has a yield slightly inferior to that of the WT. All the other *bs* lines show saccharification yields at least as good as the WT, and nine lines exceed 120%. The maximum yields reach 143.7%, 144% and 152.8% for the mutant lines 4179.1, 8469.1 and 7591.1 respectively. Error bars represent standard deviations on biological repeats.

lines 4179 and 7591 displayed the lowest lignin content. Modifications of the thioacidolysis yields and S/G ratios were also observed in these lines (Table 1), suggesting the mutations in the *bs* lines impact lignin biosynthesis both in quantity and structure. The two mutant lines 4179 and 7591 displayed the most drastic decrease in these features, indicating that these lines not only have the lowest lignin levels but the remaining lignin polymer possesses a much higher proportion of resistant interunit bonds : whereas in wild type, only 3 out of 4 units are involved in resistant bonds, up to 9 out of 10 are in the mutant lines analysed. The S/G ratio, representative of the proportions of the syringyl and guaiacyl monomers, drops from about 1.90 in the wild type to 0.66 and 0.75 in the 4179 and 7591 lines respectively. The 2 or 3 fold decrease in the proportion of S units shows that these units are the most affected by the mutation. These characteristics present themselves as the most likely explanation for the particularly increased saccharification yields in these lines: where Klason results alone could explain this increase, the majority of the remaining lignin bonds being resistant, they form more branched polymers that adsorb less to the polysaccharide chains than the  $\beta$ -O-4 linked linear polymers, and therefore leave more space for the action of digesting enzymes for saccharification. Furthermore, the lines 4179 and 7591 displayed additional thioacidolysis markers that are known to be specific to *cad* mutations (Kim, Ralph et al. 2002; Lapierre, Pilate et al. 2004) (supplementary Fig. 1). These markers originate from aldehydes incorporated in lignins, which occurs in case of CAD disruption. Since it has been shown that the accumulation of cinnamaldehyde derivatives in lignified tissues cause a brown/red color in *cad* mutants (Higuchi et al, 1994), this would explain the *bs* phenotype that first led to the identification of these two particular lines. These lines, besides the increase of up to 50% of their saccharification yields, had the particularity of showing approximately the same biomass than the wild type (supplementary Fig. 2), which makes them particularly interesting in the goal of producing biomass for biofuel production.

Table 1. Lignin analysis on extract free stems from the *bs* lines with saccharification yields over 120% of the WT. Klason analysis (%KL) is expressed in % of cell wall residue. Thioacidolysis yield is in  $\mu\text{moles/gLK}$ . The results are the mean values between repeats with their standart errors in brackets. A particularly significant reduction in lignin content can be observed in the lines 4179.1 and 7591.1. Thioacidolysis yields are also significantly low in these two lines showing a reduction in  $\beta\text{-O-4}$  bonds in the remaining lignin polymer. The important decrease of the S/G monomer ratio in these lines is representative of the severe modification of the lignin structure and the loss of S units that in also shown by the respective amounts of each monomers (%H, %G, %S). (N/A) not available

	%KL	yield (H+G+S)	S/G	%H	%G	%S
<b>WT Bd21.3</b>	<b>18 (<math>\pm</math> 0.15)</b>	<b>833 (<math>\pm</math> 47)</b>	<b>1.96 (<math>\pm</math> 0.08)</b>	<b>4.2 (<math>\pm</math> 0.1)</b>	<b>32.52 (<math>\pm</math> 0.84)</b>	<b>63.3 (<math>\pm</math> 0.9)</b>
4179.1	12,64 ( $\pm$ 0.16)	311 ( $\pm$ 11)	0.67 ( $\pm$ 0.01)	5.4 ( $\pm$ 0.25)	56.7 ( $\pm$ 0.45)	37.9 ( $\pm$ 0.35)
8469.1	17.47 ( $\pm$ 0.005)	1106 (N/A)	2.07 (N/A)	3.3 (N/A)	31.5 (N/A)	65.2 (N/A)
7591.1	13.46 ( $\pm$ 0.27)	325 ( $\pm$ 8)	0.75 ( $\pm$ 0.02)	8.4 ( $\pm$ 0.1)	52.4 ( $\pm$ 0.6)	39.2 ( $\pm$ 0.65)
7058.2	18.4 ( $\pm$ 0.095)	1093 (N/A)	2.06 (N/A)	3.5 (N/A)	31.5 (N/A)	65.0 (N/A)
3851.1	16.76 ( $\pm$ 0.005)	534 ( $\pm$ 8.5)	1.43 ( $\pm$ 0.02)	5.0 ( $\pm$ 0.3)	39.1 ( $\pm$ 0.25)	55.9 ( $\pm$ 0.35)
7058.3	19.66 ( $\pm$ 0.805)	937 (N/A)	2.04 (N/A)	4.1 (N/A)	31.5 (N/A)	64.3 (N/A)
6103.1	16.67 ( $\pm$ 0.19)	325 ( $\pm$ 29)	1.16 ( $\pm$ 0.025)	4.2 ( $\pm$ 0.2)	44.4 ( $\pm$ 0.55)	51.4 ( $\pm$ 0.55)
90.4	16.27 ( $\pm$ 0.18)	600 ( $\pm$ 16.5)	1.23 ( $\pm$ 0.015)	5.4 ( $\pm$ 0.1)	42.4 ( $\pm$ 0.2)	52.3 ( $\pm$ 0.3)
253.2	18,61 ( $\pm$ 0.21)	556 ( $\pm$ 27.5)	1.72 ( $\pm$ 0.02)	4.1 ( $\pm$ 0.1)	35.3 ( $\pm$ 0.2)	60.6 ( $\pm$ 0.03)

### **Lignin specificities of two mutant lines**

Further analysis on the *bs* lines 4179 and 7591 with the highest saccharification yields were performed. The thioacidolysis specific markers of the *cad* mutation that were found (Table 2), such as coniferaldehyde end groups, indene G 354 and indene S 384 (Kim, Ralph et al. 2002; Lapierre, Pilate et al. 2004), are byproduct monomers originating from aldehyde structures. Indene 354, representing the presence of coniferaldehyde incorporated in  $\beta$ -O-4 is totally absent in wild type lines and only present in small amounts in the mutants. In contrast, there is an over 20 fold increase in the quantity of indene 384 in the mutant lines compared to the wild type, this inden represents the amount of sinapaldehyde incorporated in the lignins.

Analysis of ester linked phenolic acids (Table 3) show a significative decrease in *p*-coumaric acid in mutants, which would be attributed to the decrease of regular S units since they are linked together in grass lignins, and only a slight decrease in ferulic acid content. Despite these results strongly suggesting a CAD deficiency, they also show that the two mutant lines described slightly differ from known maize or sorgho respective *bm* and *bmr cad* mutants in which both sinapaldehyde and coniferaldehyde are accumulated in lignins and in which the S/G ratio remains relatively stable and amounts of ferulic acid do not change.

A first explanation of the lignin composition in the described lines could come from the fact that during monolignol biosynthesis, the CAD enzymes first catalyse the coniferaldehyde to coniferyl alcohol reaction, and have a lower affinity for sinapaldehyde that would only be processed in a second time. However, this peculiar result could also be due to the presence of a redundant *CAD* gene compensating the loss of activity from the main *CAD* involved in lignification, such as in *Arabidopsis thaliana* (Sibout, Eudes et al. 2005) or rice (Li, Yang et al. 2009).

The brown/red coloration that first led to the phenotyping of these lines was observed on unstained stem cross sections (Fig. 4a, b) and must be taken into account when observing

Table 2. Lignin analysis on *bs* lines 4179 and 7591 by thioacidolysis. Results are given in  $\mu\text{mole/g}$  of Klason lignin and are the mean values between repeats with their standart errors in brackets. A significative increase in syringaldehyde end-groups incorporated in the lignin polymer can be observed, that is confirmed by the presence of this residue's derivatives inden 354 and inden 384, showing the pathway to synapil alcohol is the most affected in the mutants. A slight but hardly significative increase of Coniferaldehyde end-groups and vanillin residues corroborate this hypothesis.

	Coniferaldehyde end-groups	inden 354	inden 384	vanillin	Syringaldehyde end-groups
WT Bd21.3	0.9 ( $\pm$ 0.05)	0.0	1.8 ( $\pm$ 0.05)	3.9 ( $\pm$ 0.15)	2.1 ( $\pm$ 0.05)
4179.1	1.0 ( $\pm$ 0.1)	3.6 ( $\pm$ 0.1)	40.2 ( $\pm$ 0.3)	5.3 ( $\pm$ 0.05)	11.6 ( $\pm$ 0.45)
7591.1	1.2 ( $\pm$ 0.1)	2.8 ( $\pm$ 0.05)	46.7 ( $\pm$ 0.45)	4.5 ( $\pm$ 0.55)	12.9 ( $\pm$ 0.8)

Table 3. Analysis of cell wall bound hydroxycinnamic acids on WT Bd21-3 and *bs* lines 4179 and 7591 by thioacidolysis. Results in  $\text{mg/g}$  of CellWall Residue represent the mean values between repeats with their standart errors in brackets. The most significative decrease is observed for p-coumaric acid (CA) both ester and ether bound. Ferulic acid (FA) esters bound to cell wall arabinoxylans do not seem affected, however ether bound FA crosslinking lignins to the cell wall polysaccharides are less present. The occurrence of synaptic acid is slightly increased in the mutant lines.

	CA		FA		SA	
	ester bound	ether bound	ester bound	ether bound	ester bound	ether bound
Bd23-1	11.77 ( $\pm$ 0.28)	1.68 ( $\pm$ 0.055)	4.67 ( $\pm$ 0.045)	3.92 ( $\pm$ 0.0035)	traces	traces
4179	6.81 ( $\pm$ 0.085)	0.64 ( $\pm$ 0.05)	5.08 ( $\pm$ 0.02)	1.18 ( $\pm$ 0.09)	0.24 ( $\pm$ 0.015)	0.13 ( $\pm$ 0)
7591	8.75 ( $\pm$ 0.1)	0.98 ( $\pm$ 0.02)	4.58 ( $\pm$ 0.005)	1.90 ( $\pm$ 0.02)	0.13 ( $\pm$ 0.01)	0.26 ( $\pm$ 0.045)

stained sections. Mäule staining (Fig. 4c, d) shows a predominance of S units in interfascicular collenchyma and a predominance of G units in vascular bundles and perivascular schlerenchyma in the wild type. The same staining on the *cad* mutant lines show an almost complete disappearance of lignin in interfascicular collenchyma, but remaining G lignin in the vascular bundles and perivascular schlerenchyma. In comparison, the *Arabidopsis thaliana cadc-cadd* double mutant loses any staining with Mäule and looks white (Sibout, Eudes et al. 2005), suggesting that CAD activity is not totally disrupted in the analyzed lines and that the S units are the most affected. Phloroglucinol staining (Fig. 4e, f) in wild type also shows that the lignified tissues in *Brachypodium* stems are the vascular bundles, the perivascular schlerenchyma and the interfascicular collenchyma. This staining intensity on mutant stems is highly diminished but not completely gone, despite the important decrease in lignin quantity. This is mostly due to the important proportion of cinnamaldehydes incorporated in the remaining lignin, this staining binding with the most specificity to aldehyde residues (Kim, Ralph et al. 2002). Residual staining can particularly be observed in the vascular bundles and perivascular schlerenchyma. Stainings confirm the previous lignin analysis and suggest the hypothesis of several CAD enzymes being involved in lignifications and BdCAD1 possessing localisation specificity for interfascicular collenchyma.

### **Identification of a candidate *CAD* gene causing the brown stem phenotype**

The crossing of the *Bdcad1-1* line with a wild type Bd21 revealed at the F2 generation a Mendelian like segregation of the *cad* phenotype of the lignins. This indicated that the phenotype was most likely caused by a mutation in one locus and lead us to the candidate gene described hereafter.

Phylogenetic analysis represented in the phylogram (Fig. 5), allowed to determine *BdCAD1* as the most likely candidate gene. This analysis was performed on known or putative CAD protein

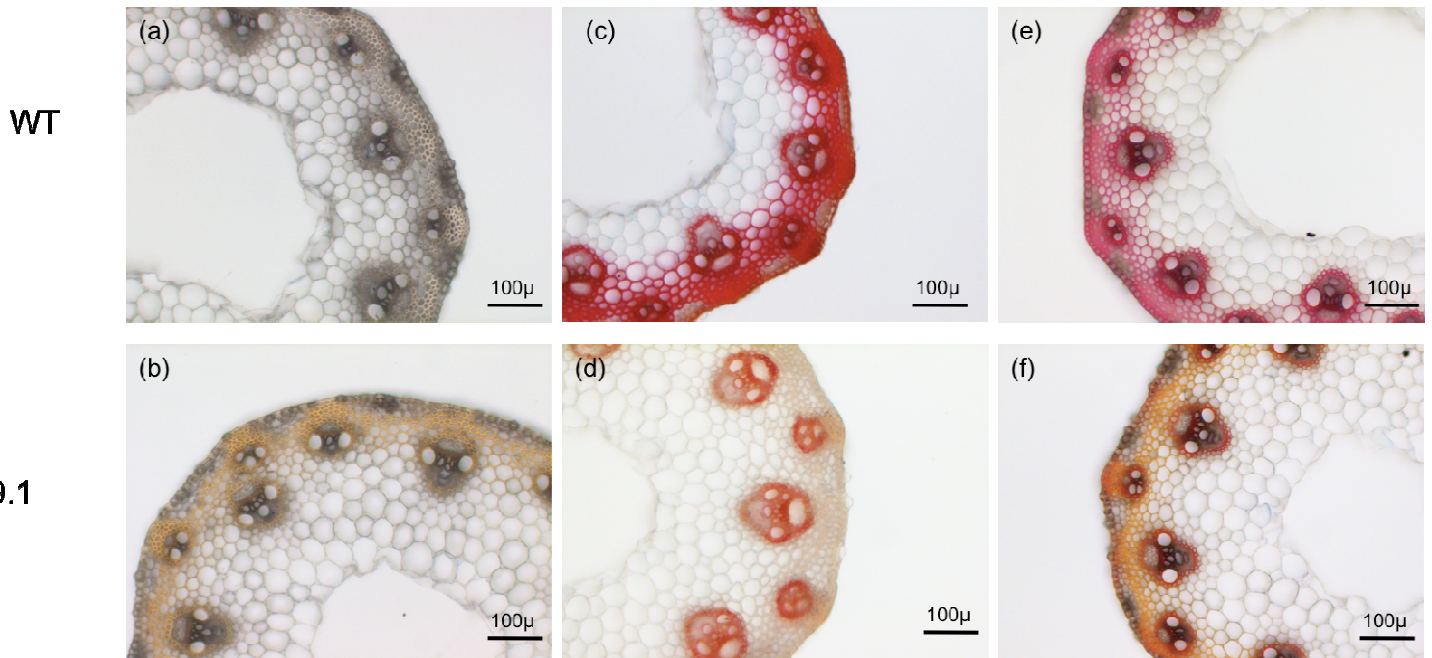


Figure 4: stem cross sections of lower internodes of wild type Bd21-3 and 4179.1 mutant. (a), (b) unstained; (c), (d) with Mäule staining ; (e), (f) Phloroglucinol staining. Both stainings are more intense in the WT. (a), (b) Mäule staining colors S units in red and G units in brown. S units are the most abundant in interfascicular collenchyma and perivascular sclerenchyma (red), lignin in vascular bundles is constituted of G units. This staining on the mutant line shows a complete loss of S lignin in the interfascicular collenchyma, and in a lesser proportion in the perivascular sclerenchyma. G units in the vascular bundles seem to remain unchanged. A loss of the natural orange color is observed, mostly due to the protocol involving a concentrated HCl wash step. (e), (f) Pink/red staining shows the presence of lignin and confirms the localization of the lignified tissues. Phloroglucinol specifically binds to aldehyde residues in the lignin polymer (Kim et al, 2001). The mutant line (f) shows very important decrease in lignin quantity, especially since the remaining colour is due to the reagent binding to the excess aldehyde residues present in the mutant lignin. As previously shown, this decrease is the most important in the interfascicular collenchyma and lignin quantity is less affected in the vascular bundles and perivascular sclerenchyma.

sequences from several species, including those of *Brachypodium distachyon* which were obtained by BLAST on the 4X then 8X genome releases (brachy.org). Seven *CAD* genes were identified in *Brachypodium* (Vogel, Garvin et al. 2010), that will be hereafter referred to as *BdCAD1* to 7 accordingly to their positions on the chromosomes. *CAD* genes with a major role in main lignified tissues represent the Class I described by Barakat (Barakat, Bagniewska-Zadworna et al. 2009) in which *BdCAD1* can be found. The *BdCAD1* protein has 73% identity and 85% similarity with *AtCAD-C*, 74% identity and 85% similarity with *AtCAD-D* (Sibout, Eudes et al. 2005), 89% identity and 96% similarity with *OsCAD2* (Tobias and Chow 2005) and 89% identity and 95% similarity with *ZmCAD2* (Halpin, Holt et al. 1998) that all belong to the Class I and are involved in lignification. Furthermore, the *BdCAD1* protein possesses the conserved residues constituting the substrate binding sites, the glycine rich domain GXG(X)<sub>2</sub>G at the NADPH binding site, as well as the residues involved in the binding of the catalytic and structural zinc ions. In addition all residues that are highly conserved in plant *CAD* proteins are found in *BdCAD1*. These residues are identified on the protein sequence shown in supplementary data (Supplementary Fig. 3).

Sequencing of *BdCAD1* in the two mutant lines revealed missense mutations in their coding sequences. The lines 4179 and 7591 will henceforth be respectively referred as *Bdcad1-1* and *Bdcad1-2*. *In silico* predictions of the 3 dimensional structures of the mutant *Bdcad1* (Fig. 6a, b, c, d) suggest these mutations could cause a decrease in the protein activity. The mutation in the line *Bdcad1-1* is located at G192D, in the middle of the conserved glycine residues at the co-factor binding site. Even though this amino acid is probably not directly involved in the NADPH fixation, it is highly conserved throughout the *CAD* family. Moreover, the modification in size and charge generated by the replacement of a glycine by an aspartic acid at a location of importance for the binding of both the substrate and the co-factor can create a disturbance in their fixation. In particular, 3D surface modelization (Fig. 6 c,d) shows the



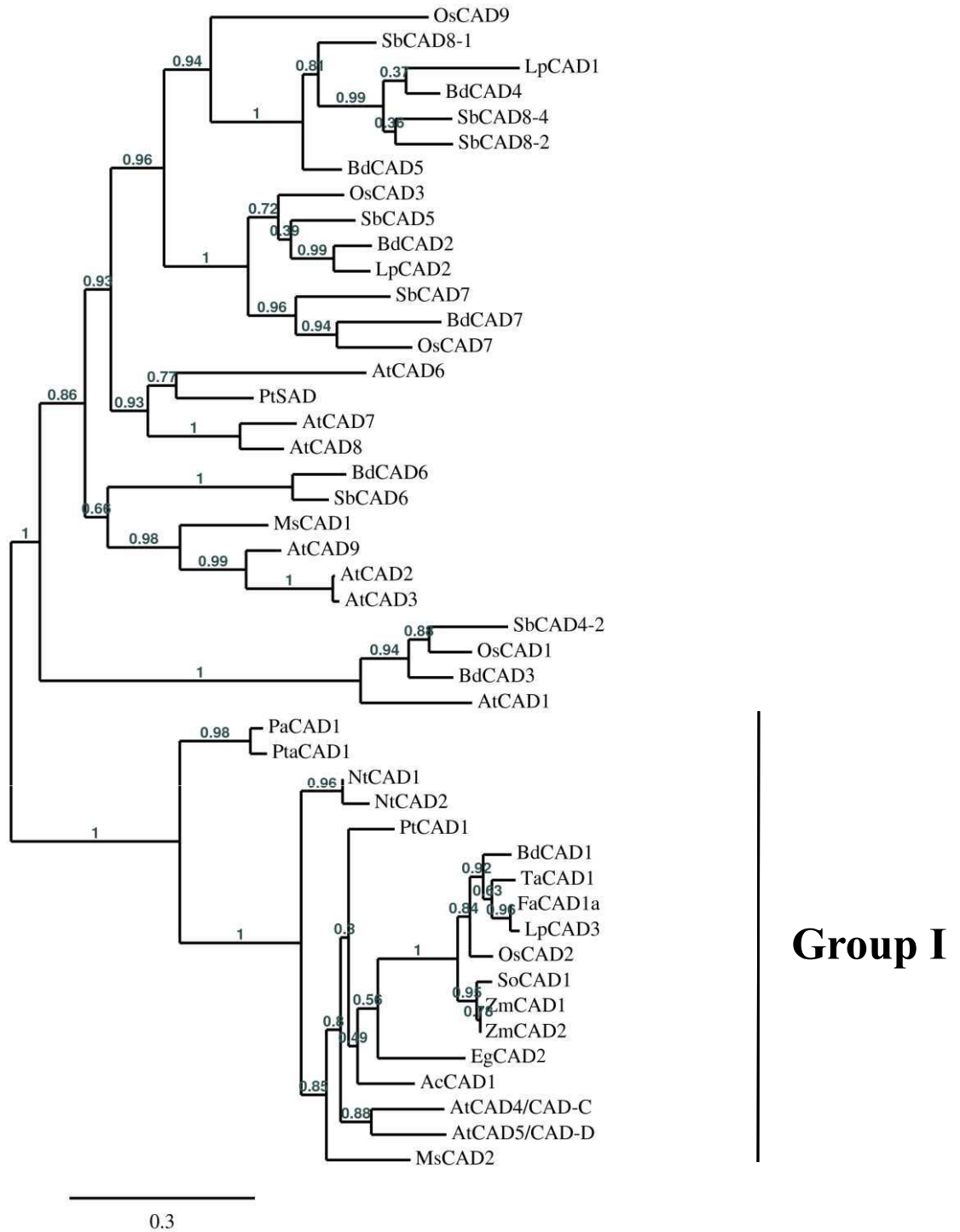


Figure 5. Phylogenetic analysis of CAD proteins among plants, including translated *Brachypodium* putative *CAD* genes. Proteins that have been shown to be involved in lignification belong to the Group I (according to the work of Ma et al, 2010) into which BDCAD1 clusters. Numbers in grey show the bootstrap value associated with each branch of the phylogram.

obstruction of a pocket in the protein, presumably involved in the binding of the substrate and the cofactor. This pocket also gives accesses to the fixation site of the catalytic zinc and this obstruction might prevent the metal ion to bind to the protein.

BdCAD1 in the *Bdcad1-2* line carries two missense mutations G99V and S286F. The first mutation is close to an amino acid involved in the binding of the structural zinc, but according to the 3D structure prediction and the nature of the amino acids in question, it does not appear to have an impact either on the structure or the function of the protein. However, the second mutation is located next to a substrate binding site at the extremity of one of the sub-units that is involved in the catalytic activity of the second sub-unit of the CAD dimer. The difference of charge and size between Phenylalanine and Serine at this site suggests this mutation can interfere with the proper activity of the enzyme (Fig. 5e, f, g, h).

Moreover, the existence of several mutants in the same gene *BdCAD1*, originating from independent mutagenesis events and showing the same known *cad* lignin phenotypes is an additional indication that the mutations in our candidate gene are the cause of that phenotype.

### **Expression pattern and activity of BdCAD1 are consistent with lignification**

RNA was extracted from Brachypodium stems (nodes, internodes and leaf sheath), retro-transcribed at four growth stages and quantitative RT-PCR was performed using primers specific to the 3' end of each *BdCAD* genes. Results (Fig. 7a) indicated that BdCAD1 is the most expressed gene among the CAD family. A slight but significant increase of expression could be observed until the flowering stage at 9 weeks, after panicle exertion but previous to anthesis, then a decrease in maturing plants at 12 weeks.

Semi quantitative RT-PCR was carried out on separate tissues of the main stem at the 9 weeks stage (nodes, young internodes, old internodes, leaves, leaf sheaths, peduncle and spikelets) to

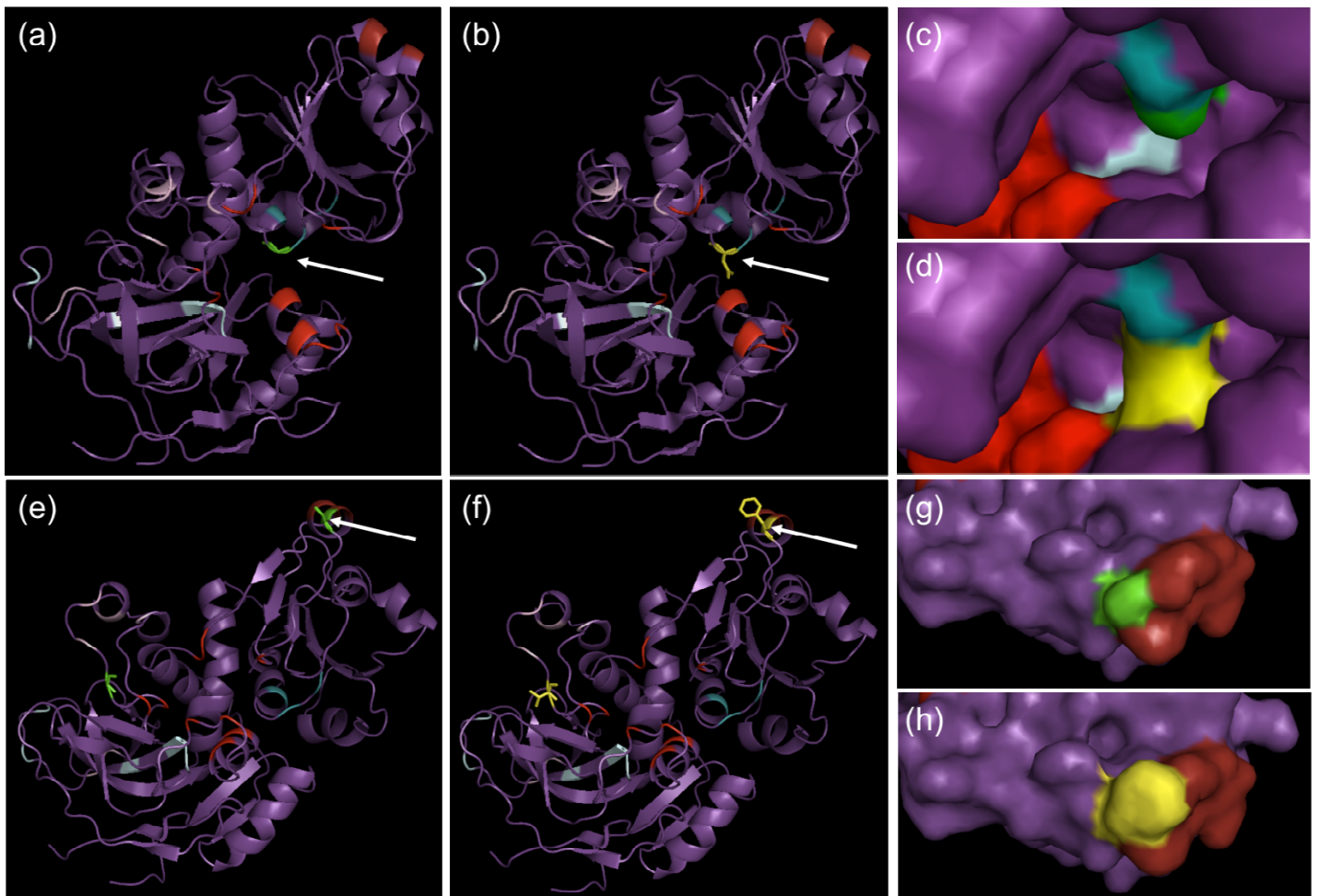


Figure 6. Three dimensional structure predictions of BdCAD1. (a) wild type Bd21-3 compared to (b) *Bdcad1-1* in cartoon view and close ups on the substrate pocket in surface view of WT (c) compared to *Bdcad1-1* (d). (e) wild type compared to (f) *Bdcad1-2* in cartoon view and close ups on the amino acids involved in the active site from the second monomer of wild type (g) and *Bdcad1-2* (h). Red: active site; teal: cofactor binding site; light blue: catalytic Zn binding site; light blue: structural binding site; green: wild type amino acid; yellow: mutant amino acid.

identify the major localization of the expression of *BdCAD1*, and results showed it was the most expressed in young developing internodes and in the peduncle.

These results are strongly associated with genes involved in lignification and confirm the choice of *BdCAD1* as the main candidate for cinnamyl alcohol dehydrogenase of monolignol biosynthesis.

Activity assays of wild type Bd21.3 (Fig. 7b, c) indicate a correlation between the expression of *BdCAD1* and CAD activity in *Brachypodium* stems during development. CAD activity increases until the 9 weeks stage then decreases. In both the *Bdcad1-1* and *Bdcad1-2* lines at the 9 weeks stage, a loss of up to 20% in CAD activity was observed. This result might not be as strong as could have been expected, however it can be explained by the lack of specificity of the assay. Indeed, as it measures total CAD activity in the whole stem, it does not reflect any tissue specificity, nor does it discriminate between the activity of BdCAD1 from that of the six other BdCAD enzymes. However, the detected decrease in CAD activity is significant and similar results have been observed in *cad* mutants from other species such as the gold hull rice mutant (Zhang, Qian et al. 2006) or in barley (Claire Halpin, personal communication). The remaining activity can be due to other CAD enzymes, presumably with various tissue specificity and different biological functions, especially since we have observed the increase in expression of *BdCAD2* and *BdCAD3* in the *Bdcad1-1* line at this growth stage (data not shown).

### **Functional complementation of *cad* mutants**

The coding sequence of *BdCAD1* was cloned under control of the CaMV35S promoter into the Gateway system PGWB2 that was used to transform the *Arabidopsis thaliana cadc-cadd* double mutant in the WS accession (Sibout, Eudes et al. 2005). As early as the T1 transformed lines, the brown color of the *cadc-cadd* stems was gone in the stems of the lines carrying and

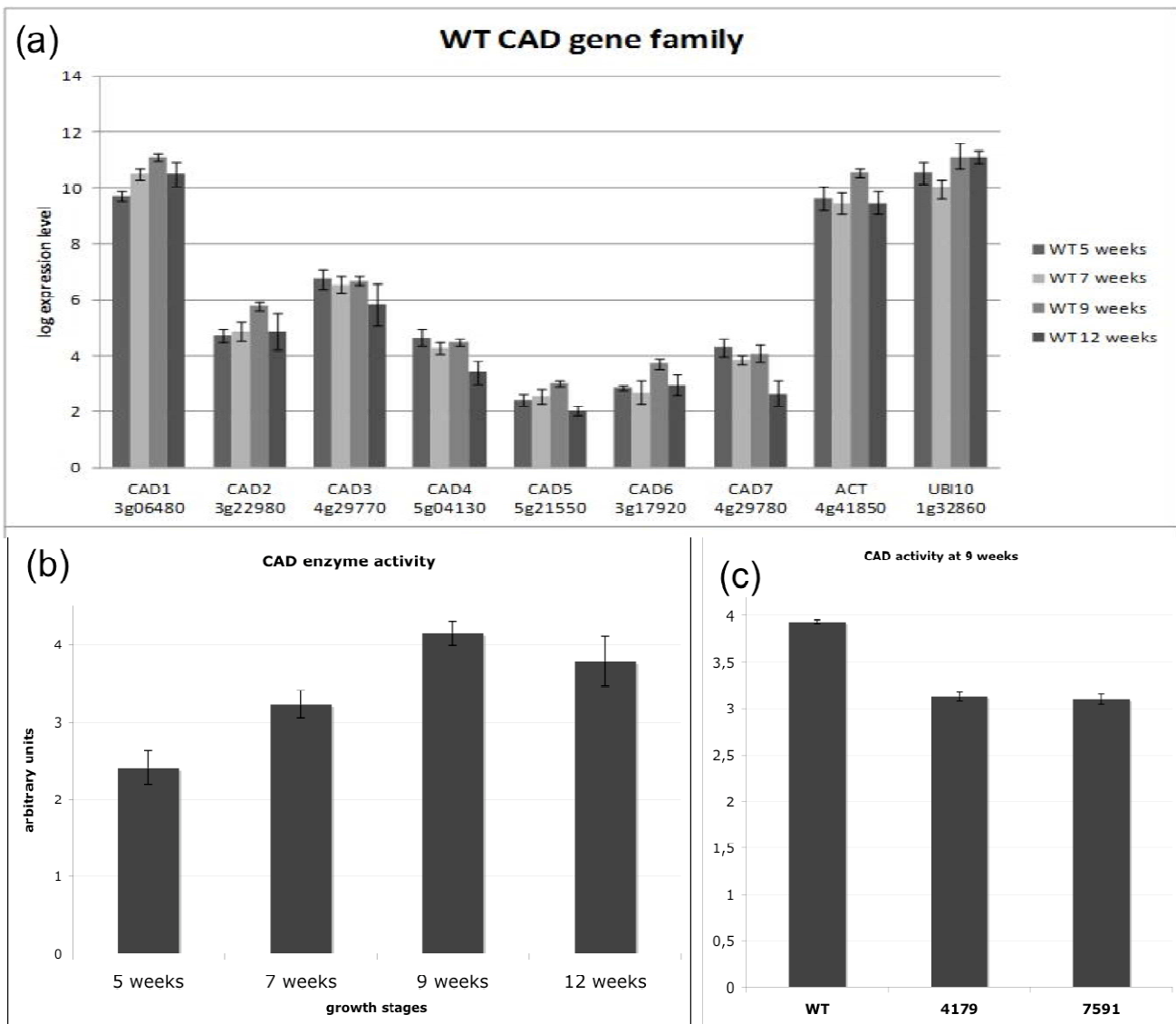


Figure 7. (a) Quantitative RT-PCR on Bd21.3 at 4 growth stages on complete stems (internodes + nodes + leaf sheath) for all seven annotated *BdCAD* genes. (3 biological repeats for each stage) 5 weeks: end of the vegetative stage; 7 weeks: previous to emergence of florets; 9 weeks: shortly after emergence of the florets; 12 weeks: just prior to start of dessication. The significantly most expressed gene is *BdCAD1*, showing an increase in expression during the three first stages, then a decrease in the last stage. This behaviour is consistent with known lignin genes, especially *CAD* in other species. (b) *CAD* activity assays on complete stems of wild type Bd21-3 at four growth stages and (c) comparison at 9 weeks between the WT and the two mutant lines 4179 and 7591. Units are arbitrary and represent the director coefficient of the linear correlation from the activity kinetics. (a) Correlated to the gene expression levels shown previously, the total *CAD* activity in the stems rises during the first three stages, then decreases at 12 weeks. (b) At 9 weeks, *CAD* activity levels of the mutant lines were significantly lower than in the WT, showing a decrease of over 20%.

expressing the insert. Their color, identical as that of the wild type, showed the absence of accumulation of aldehyde precursors in lignin (supplementary data Fig. 4). In the T3 lines containing a single copy of the insert at a homozygous state that were analysed, Klason lignin content was recovered at between 99.9% and 102% of the WT content, when the content in the *cadc-cadd* double mutant is only around 78% of the WT (Table 4). In addition, the quantity of acido-soluble lignins that was increased in the double mutant was back to values close to the WT in the lines analyzed. The S/G ratio that had dropped to 0.05 in *cadc-cadd* mutants compared to the 0.36 of the WT is restored at between 0.37 and 0.4 in the analyzed complemented lines. Although the *BdCAD1* gene is expressed under a strong constitutive promoter, the lignin levels do not exceed those of the wild type, probably due to the limiting factor of the preceding steps in the monolignol pathway. It was shown that loss of function of CAD activity in the *cadc-cadd* mutant does not impact the transcript levels of other genes of the pathway by feedback (Sibout, Eudes et al. 2005) and the stable lignin content suggest that more CAD expression does not apply a positive feedback either, that would increase the production of precursors. Moreover, presence of inden 354 and inden 384, the thioacidolysis diagnostic markers for *cad* mutation, are down to traces corresponding to the levels found in WS wild types. These results indicating the recovery of quantity and composition of lignins at the wild type level in the complemented lines confirm that *BdCAD1* has a cinnamyl alcohol activity involved in lignification.

The coding sequence of *BdCAD1* was also cloned under control of the *ZmUBI* promoter in a pIJPB construct that was used to transform embryogenous callus of the *Bdcad1-1* mutant line.

The lines carrying the construct showed no sign of *bs* phenotype (Supplementary Fig. 5). Lignin analysis in the complemented lines (Table 5) revealed an increase in thioacidolysis yields back to the levels of the WT Bd21-3, even exceeding it in some lines, showing an increase in the presence of  $\beta$ -O-4 bound units compared to the mutant line. The S/G ratio, that

Table 4. Lignin analysis on T3 *Atcadc-cadd* lines complemented by *BdCAD1*. WT: Wild Type Bd21-3; *cadc cadd*: *Arabidopsis thaliana cadc cadd* double mutant; CADcp: *Arabidopsis thaliana cadc cadd* double mutant complemented with *BdCAD1*. %CWR : proportion of Cell Wall Residue left after extraction. KL: Klason Lignin in %CWR; ASL: Acido Soluble Lignin; H+G+S: thioacidolysis yield of H, G and S monomers in  $\mu\text{mole/gLK}$ ; S/G: ratio of  $\beta$ -O-4 bound S units to  $\beta$ -O-4 bound G units; Cald end groups: Coniferaldehyde incorporated in the lignin monomer; IndensG 354, IndeneS 384, thioacidolysis markers of CAD mutation. Results are the mean values between repeats with their standart errors in brackets Tr: traces. For all lignin characteristics analyzed, the levels in the complemented lines are back to WT levels.

	% CWR	KL (% CWR)	ASL (% CWR)	S/G	%H	%G	%S	Thioacidolysis yields (in $\mu\text{mole/gLK}$ )			
								H+G+S	Cald end- groups	IndeneG 354	IndeneS 384
WT	63.3 ( $\pm 0.3$ )	19.79 ( $\pm 0.095$ )	1.40 ( $\pm 0.015$ )	0.36 ( $\pm 0$ )	0.5 ( $\pm 0$ )	73.2 ( $\pm 0.05$ )	26.4 ( $\pm 0.05$ )	1663 ( $\pm 9.5$ )	4.9 ( $\pm 0.05$ )	Tr	Tr
<i>cadc cadd</i>	52.5 ( $\pm 0$ )	15.53 ( $\pm 0$ )	1.88 ( $\pm 0.02$ )	0.05	1.1	94.3	4.5	77	14.4	68.2	36.2
CADcp 1- 5	63.3 ( $\pm 0.9$ )	19.78 ( $\pm 0.35$ )	1.34 ( $\pm 0.005$ )	0.40 ( $\pm 0$ )	0.6 ( $\pm 0$ )	71.3 ( $\pm 0.15$ )	28.2 ( $\pm 0.1$ )	1668 ( $\pm 3$ )	5.7 ( $\pm 0.15$ )	Tr	Tr
CADcp 11-5	62.5 ( $\pm 1.35$ )	19.83 ( $\pm 0$ )	1.39 ( $\pm 0.005$ )	0.37 ( $\pm 0$ )	0.7 ( $\pm 0$ )	72.4 ( $\pm 0.1$ )	27.0 ( $\pm 0.15$ )	1588 ( $\pm 19.5$ )	5.1 ( $\pm 0.05$ )	Tr	2.3 ( $\pm 0.1$ )
CADcp 22-4	68.1 ( $\pm 0.1$ )	20.22 ( $\pm 0$ )	1.36 ( $\pm 0.01$ )	0.37 ( $\pm 0$ )	0.5 ( $\pm 0.05$ )	72.4 ( $\pm 0.05$ )	27.1 ( $\pm 0$ )	1524 ( $\pm 26$ )	3.6 ( $\pm 0.4$ )	Tr	Tr
CADcp 22-6	60.4 ( $\pm 0.55$ )	19.40 ( $\pm 0.03$ )	1.47 ( $\pm 0$ )	0.36 ( $\pm 0.005$ )	0.5 ( $\pm 0$ )	73.2 ( $\pm 0.35$ )	26.4 ( $\pm 0.35$ )	1487 ( $\pm 23.5$ )	5.3 ( $\pm 0.4$ )	Tr	7.6 ( $\pm 1.2$ )

was significantly reduced in the mutant line, was reestablished at WT levels. Mainly, the strong decrease of the diagnostic markers levels show that the mutation in *BdCAD1* can be complemented. This strongly suggests that the mutation in *BdCAD1* in the *Bdcad1-1* line was the cause of the *bs* phenotype observed and that it can be restored by the introduction of a wild type allele.

## **Conclusion**

### **Stem coloration can be an efficient phenotyping strategy for modified cell wall composition**

Stem and midrib coloration have been described in maize and sorgho as results of lignin modifications (Vignols, Rigau et al. 1995; Halpin, Holt et al. 1998; Saballos, Ejeta et al. 2009). Screening a mutant collection based on this phenotype revealed to be efficient since the great majority of these lines showed lignin modifications and improved saccharification. Moreover, thioacidolysis allowed the determination of very specific mutations according to diagnostic markers. The two *cad* mutants found displayed significant phenotypic similarities with known *cad* mutants in other species. The lignin compositions were however slightly different in some particular aspects such as the accumulation of ferulic acid.

### **BdCAD1 is the main CAD involved in the monolignol biosynthetic pathway**

The CAD family consists of 7 members in *Brachypodium distachyon*. It has been shown in several species that although one CAD isoform is the main actor of this pathway, at least one other member is involved and can partially compensate the disruption of the first enzyme (Sibout, Eudes et al. 2005; Li, Yang et al. 2009).

The activity of *BdCAD1* can be disrupted by point mutations that leads to the loss of up to 30% of the total CAD activity in the plant (activity of all combined CAD enzymes in all combined



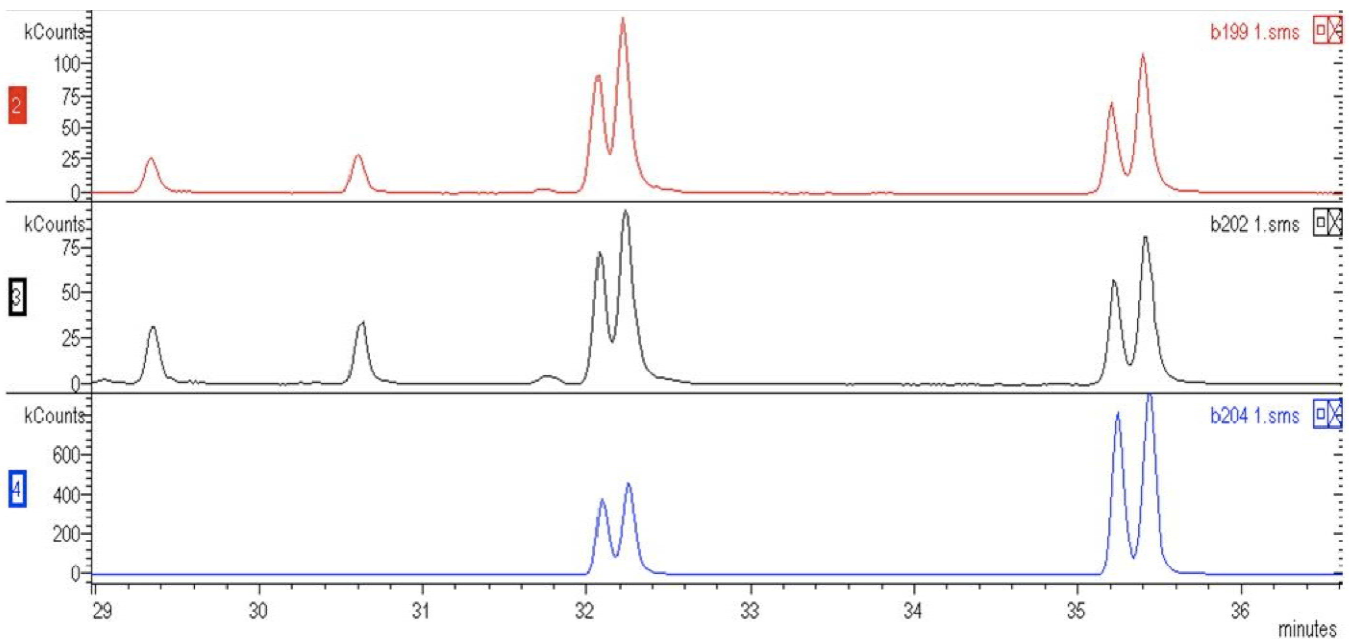
Table 5. Lignin analysis on T1 4179 lines transformed with *BdCAD1*. Thioacidolysis yields are in  $\mu\text{mole/g}$  material. P-coumaric acid (CA) and Ferulic acid (FA) are in  $\text{mg/g}$  of cell wall residue. For all lignin characteristics analyzed, the levels in the complemented lines are back to that of the WT except in line 4179Cp 5-2.

	Thioacidolysis yields		S/G	PC	Fe
	H+G+S	indene 384			
WT Bd21-3	139	0.1	1.49	9.23	7.10
4179	45	7.5	0.51	6.43	5.43
4179 Cp 1-2	145 ( $\pm 0.5$ )	0.1	1.66	9.57 ( $\pm 0.095$ )	7.33 ( $\pm 0.02$ )
4179 Cp 4-4	141 ( $\pm 2.5$ )	0.1	1.56 ( $\pm 0.01$ )	10.10 ( $\pm 0.02$ )	7.46 ( $\pm 0.09$ )
4179 Cp 7-1	169 ( $\pm 3.5$ )	0.1	1.49 ( $\pm 0.01$ )	8.94 ( $\pm 0.155$ )	8.44 ( $\pm 0.115$ )
4179 Cp 10	131 ( $\pm 2.5$ )	0.1	1.30 ( $\pm 0.005$ )	8.90 ( $\pm 0.06$ )	7.68 ( $\pm 0.13$ )
4179 Cp 5-2	33 ( $\pm 1$ )	3.6 ( $\pm 0.05$ )	0.56 ( $\pm 0.03$ )	7.80 ( $\pm 0.035$ )	5.60 ( $\pm 0.01$ )

tissues a given growth stage), which shows that BdCAD1 is involved in at least a third of all CAD activity. Histological evidence suggested that BdCAD1 is quite specific to interfascicular collenchyma, a tissue type containing mostly S lignin. This tissue specificity could explain impact particularly important on S units of the mutation described. In addition to BdCAD1, one or several other CAD enzymes could be responsible for monolignol biosynthesis in vascular bundles and perivascular sclerenchyma, tissues that contain mostly G units. In-vitro analyses of substrates could bring answers to these questions, and a BdCAD1<sub>prom</sub>::GUS construct is currently being transformed into *Brachypodium* wild type lines to verify our tissue specificity hypothesis. An additional explanation of the lignin composition in the described lines could come from the fact that during monolignol biosynthesis, the CAD enzymes first catalyse the coniferaldehyde to coniferyl alcohol reaction, and have a lower affinity for sinapaldehyde that would only be processed in a second time.

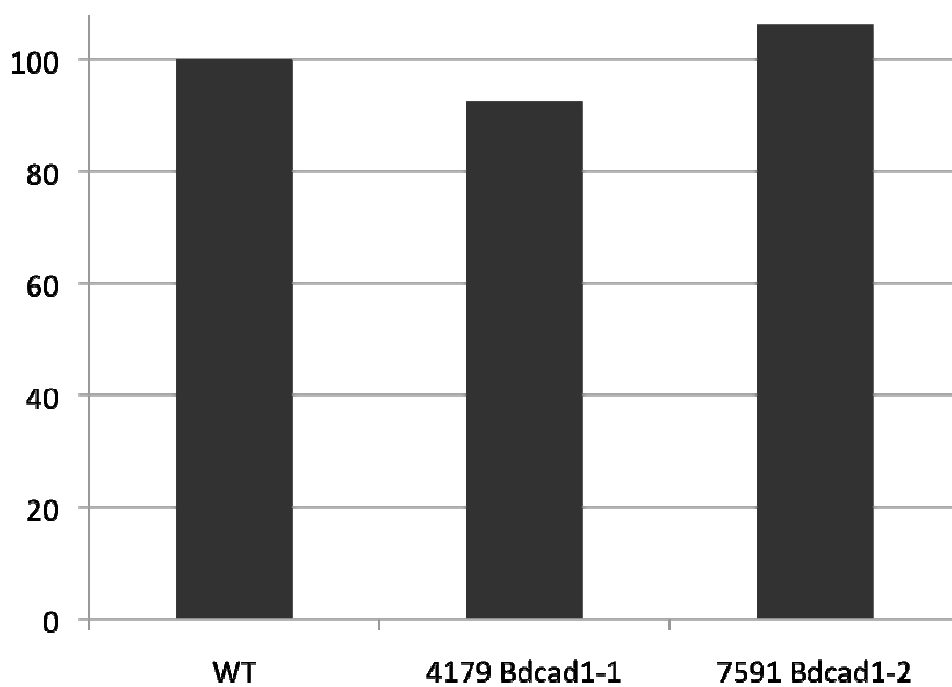
## **Aknowledgements**

This work was funded by the RENEWALL European project. We thank Joseph Tran for his insight on 3D protein modelization and Patrick Grillot, Fabrice Petitpas and Philippe Maréchal for taking care of the plants in the greenhouses and culture chambers.



Supplementary figure 1. Thioacidolysis GC-MS profiles the lignins from the 7591 (on top), 4179 (middle) and WT (bottom) lines. Area beneath the peaks show the quantities of each compound. The peaks between 35 and 36 minutes represent the S lignin units, the peaks at 32 minutes represent the G lignin units. Both have larger areas for the WT line compared to the mutants. The mutant lines profiles display two peaks, at 29.2 minutes and 30.6 minutes, that do not appear on the WT profile. These peaks represent derivatives from aldehyde accumulation in lignin, that are specifically linked to *cad* mutants.

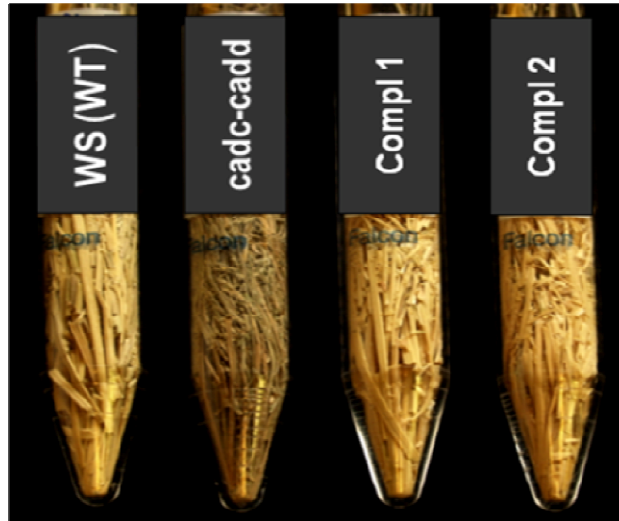
## Biomass in % of the WT



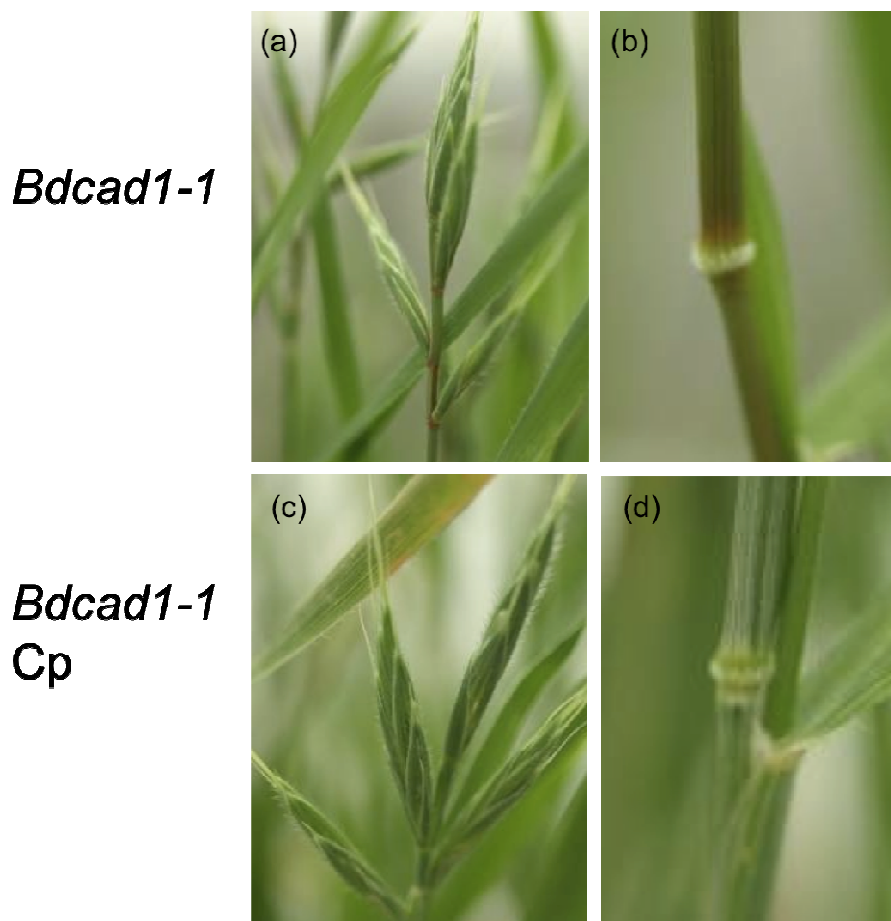
Supplementary figure 2. Biomass of plants from the two mutant lines 4179 and 7591. Five plants from each line were weighed and the average is given in % of the WT. Compared to the WT line, the 7591 line displays slightly superior biomass (106%) and 4179 displays slightly inferior (94%) biomass.



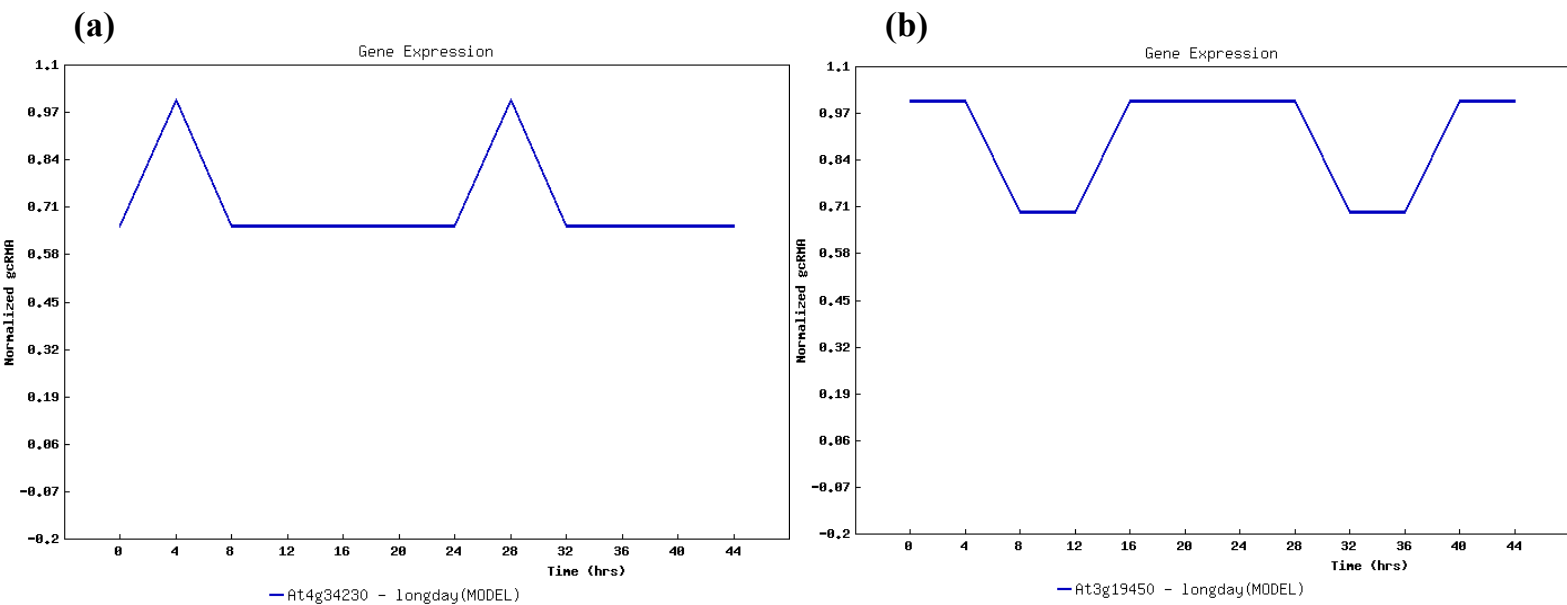
Supplementary figure 3. The sequences of BdCAD1 in the WT Bd21-3 as well as in the mutant lines 4179 (*Bdcad1-1*) and 7591 (*Bdcad1-2*) are shown. Red residues are the conserved residues constituting the substrate binding site, brown residues are part of the substrate binding site from the second monomer involved in the CAD dimer, teal residues are the cofactor binding site, blue-grey residues are involved in the binding of the catalytic zinc, and the pink residues are involved in the binding of the structural zinc. The mutation in *Bdcad1-1* occurs in the middle of the cofactor binding site and the second mutation of *Bdcad1-2* occurs next to a substrate binding site from the dimer.



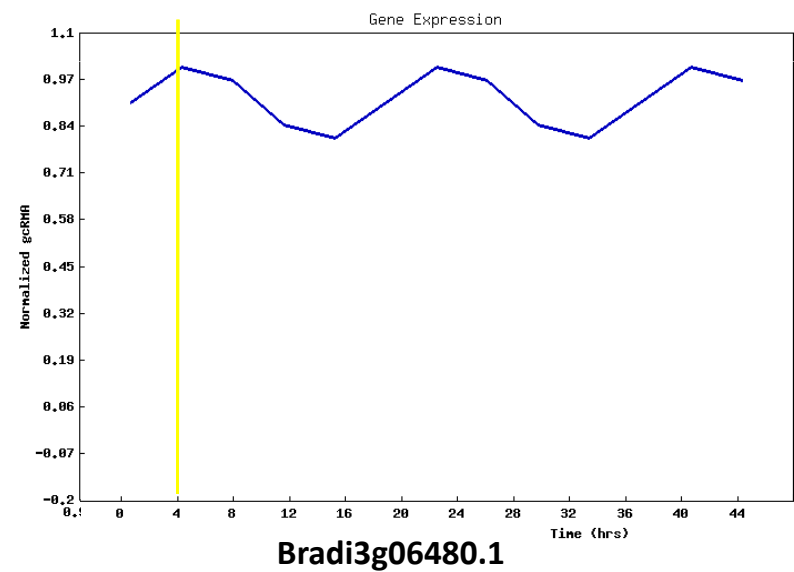
Supplementary figure 4. Arabidopsis stems shown in 15mL tubes. The *cadc-cadd* mutant stems display a brown coloration that is not observed in the WT nor in the complemented lines.



Supplementary figure 5. (a) (c) spikelets and (b) (d) from a *Bdcad1-1* mutant line and a *Bdcad1-1* line complemented with the *BdCAD1* wild type allele. The red coloration at the spikelet bases and around the nodes of the mutant line are no longer observed in the complemented line.



**Figure 1.** Niveaux d'expression en fonction du cycle circadien des gènes *AtCAD-C* (a) et *AtCAD-D* (b) dans les conditions « long day », correspondant aux conditions de culture d'*Arabidopsis thaliana* à Versailles. Le temps en heures est indiqué en abscisse, et une unité arbitraire normalisée est indiquée en ordonnée. Les deux gènes ont en commun un pic d'expression à 4h après la mise en marche des lumières.



**Figure 2.** Niveau d'expression en fonction du cycle circadien de *BdCAD1* dans les conditions de culture LDHC, les plus similaires à celles de Versailles. Le temps en heures est indiqué en abscisse, et une unité arbitraire normalisée est indiquée en ordonnée. Les barres jaunes indiquent quatre heures après le début de l'exposition à la lumière. Un pic d'expression est observé à 4h après la mise en marche des lumières.

## Annexe au chapitre II :

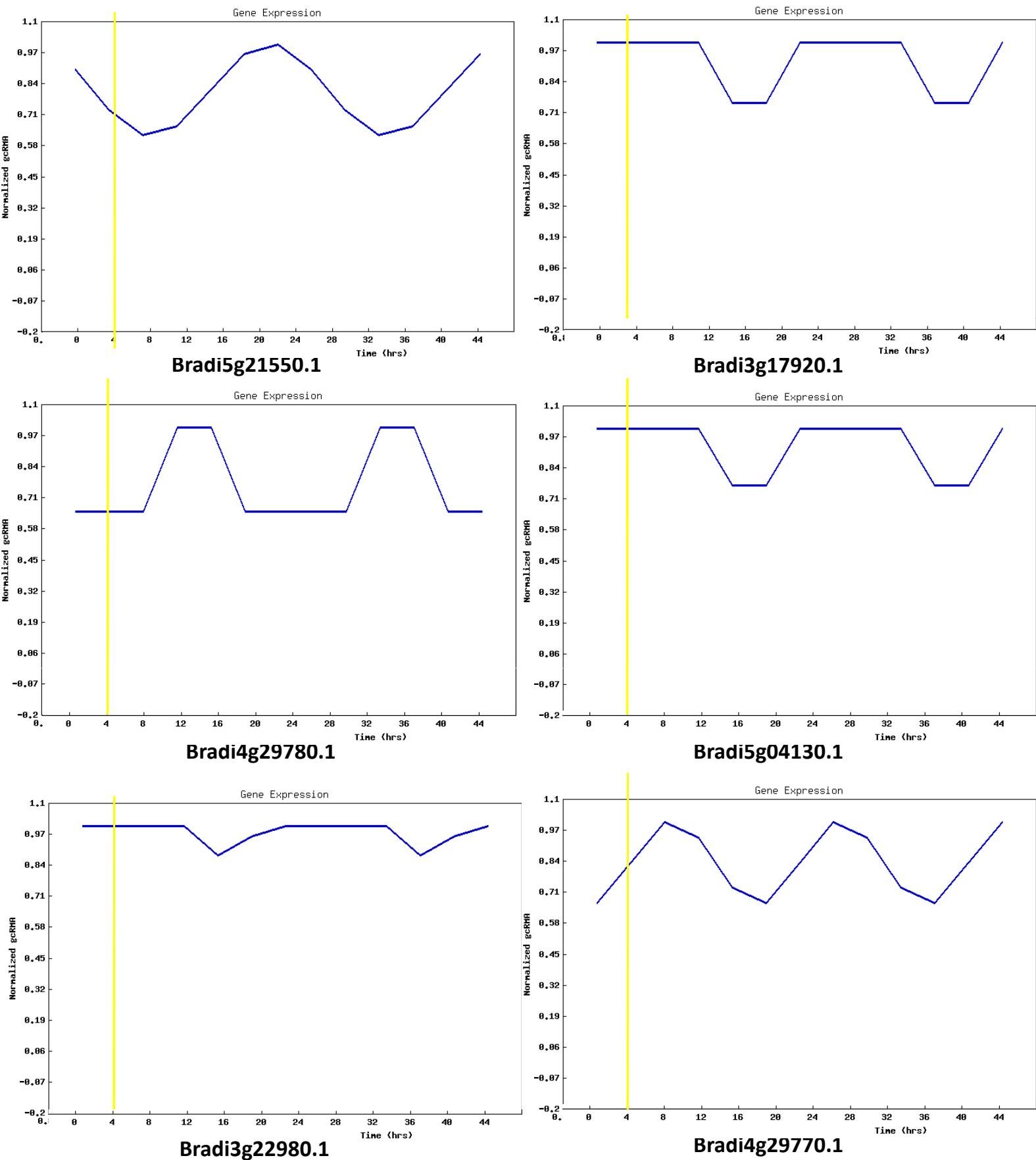
Le choix du gène candidat décrit précédemment a été appuyé par des analyses complémentaires dont les résultats sont décrits dans cette annexe.

### 1. Expression en fonction du cycle circadien

Les données ont été obtenues sur le site Diurnal (Mockler, Michael et al. 2007), à l'URL suivante : [diurnal.cgrb.oregonstate.edu](http://diurnal.cgrb.oregonstate.edu)

L'expression des gènes impliqués dans la voie de biosynthèse des monolignols varie au cours d'un cycle circadien. Les données présentes au sujet de la plante modèle *Arabidopsis thaliana* suggèrent que les deux gènes *AtCAD-C* et *AtCAD-D* ont tous les deux un pic d'expression commun au bout de quatre heures après le début du jour, ou de la mise en route des lumières en salles de culture dans les conditions de culture de Versailles (Fig. 1). Ce pic est retrouvé chez tous les gènes impliqués dans cette voie de biosynthèse, et également chez les laccases impliquées dans leur polymérisation au niveau de la paroi végétale. Aucune donnée sur le gène *CAD* majeur du riz (*OsO2t0187800*) n'existait sur ce site, c'est pourquoi le cycle d'*Arabidopsis* a été choisi comme modèle de référence pour le cycle circadien. Ainsi, tous les prélèvements concernant l'analyse des transcrits ou des protéines *CAD* ont été effectués 4h après la mise en route des éclairages afin d'en obtenir des quantités maximales. Lors de la détermination du gène candidat *CAD* de *Brachypodium*, la banque de données diurnal ne contenait pas d'information sur notre plante modèle. Ces données sont à présent disponibles et l'hypothèse de travail a pu être vérifiée : d'après les données Diurnal, le gène *BdCAD1* (*Bradi3g06480*) présente bien un pic d'expression après quatre heures d'exposition à la lumière dans les conditions de culture les plus similaires à celles de *Brachypodium* à Versailles (Fig. 2). La





**Figure 3.** Niveaux d'expression en fonction du cycle circadien de tous les gènes annotés *BdCAD* dans les conditions de culture LDHC, les plus similaires à celles de Versailles. Le temps en heures est indiqué en abscisse, et une unité arbitraire normalisée est indiquée en ordonnée. Les barres jaunes indiquent quatre heures après le début de l'exposition à la lumière.

validation de notre gène candidat ayant été faite par ailleurs, ces données permettent de renforcer l'hypothèse selon laquelle le profil d'expression au cours du temps est similaire entre *Arabidopsis* et *Brachypodium*, et nous permet de déterminer si d'autres gènes de la famille des cinnamyl alcool déshydrogénases ou bien si les gènes candidats des autres familles de la voie de biosynthèse des monolignols ont bien un profil « lignine » (Fig. 3).

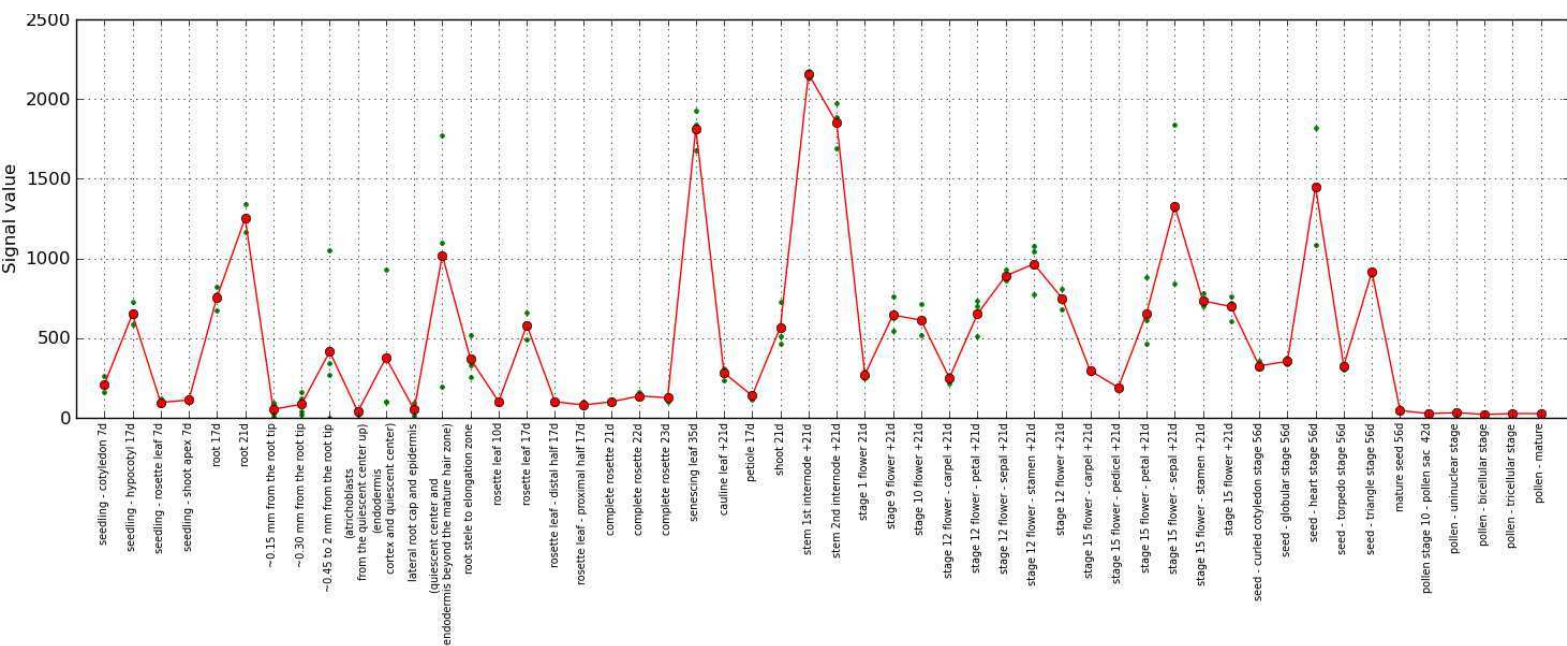
## 2. Profils d'expression en fonction des tissus

Les données ont été obtenues sur le site Bradinet, à l'URL suivante : <http://aranet.mpimp-golm.mpg.de/bradinet/>. Cet outil regroupe les données d'expression et de co-expression obtenues sur puces à ARN.

Les niveaux d'expression maximum des gènes impliqués dans la voie de biosynthèse des monolignols sont attendus dans les tissus les plus lignifiés, c'est-à-dire en particulier les entrenœuds chez les céréales. Par exemple, chez *Arabidopsis thaliana*, la hampe florale est le tissu le plus lignifié et l'on y observe les niveaux d'expression les plus importants du gène *AtCAD-D* (Fig. 4). Parmi tous les gènes *CAD* annotés chez *Brachypodium*, le candidat principal pour la voie de biosynthèse des monolignols sera celui (ou ceux) présentant des niveaux d'expression les plus élevés dans les tiges ou entrenœuds. Par ailleurs, les gènes impliqués dans la lignification doivent être coexprimés avec les gènes de formation de la paroi végétale secondaire.

D'après les données d'expression (Fig. 5), *BdCAD1* (Bradi3g06480.1) présente les niveaux d'expression les plus élevés parmi tous les gènes *CAD* de *Brachypodium*. Par ailleurs, ces niveaux d'expression sont particulièrement forts trois jours après épiaison dans le mésocotyle, les entrenœuds, le jeune épi et les nœuds, ce qui confirme le profil « lignine ».

L'analyse de coexpression montre que *BdCAD1* appartient au cluster 94. Celui-ci contenant également des gènes impliqués dans la formation de la paroi (tels que des gènes impliqués dans



**Figure 4.** Niveaux d'expression en unités arbitraires du gène *AtCAD-D* dans différents des organes indiqués en abscisse. Les niveaux les plus forts sont localisés dans les tissus à la lignification la plus importante: les hampes florales.

la biosynthèse de la cellulose, des protéines pariétales, d'autres gènes de la voie de biosynthèse des monolignols (*PAL*, *HTC*) ou impliqués dans le métabolisme secondaire).

L'hypothèse qui avait été formulée au sujet d'un second gène *CAD* pouvant être impliqué dans la voie de biosynthèse des monolignols est à nouveau suggérée : Le profil d'expression de *BdCAD4*, bien que moins fort que celui de *BdCAD1*, lui est très similaire : les tissus dans lesquels *BdCAD4* est exprimé sont le mésocotyle, l'épi, les *entre-nœuds* et les nœuds à trois jours après épiaison. Par ailleurs, il appartient au cluster 61 qui contient d'autres gènes de biosynthèse des lignines (notamment le gène *PAL* appartenant également au cluster 94).

En comparaison, les clusters des autres *CAD* annotées ne contiennent que très peu, ou pas de composés liés à la déposition de lignine (Tableau 1) :

132 : isoprenoïdes, brassinostéroïdes

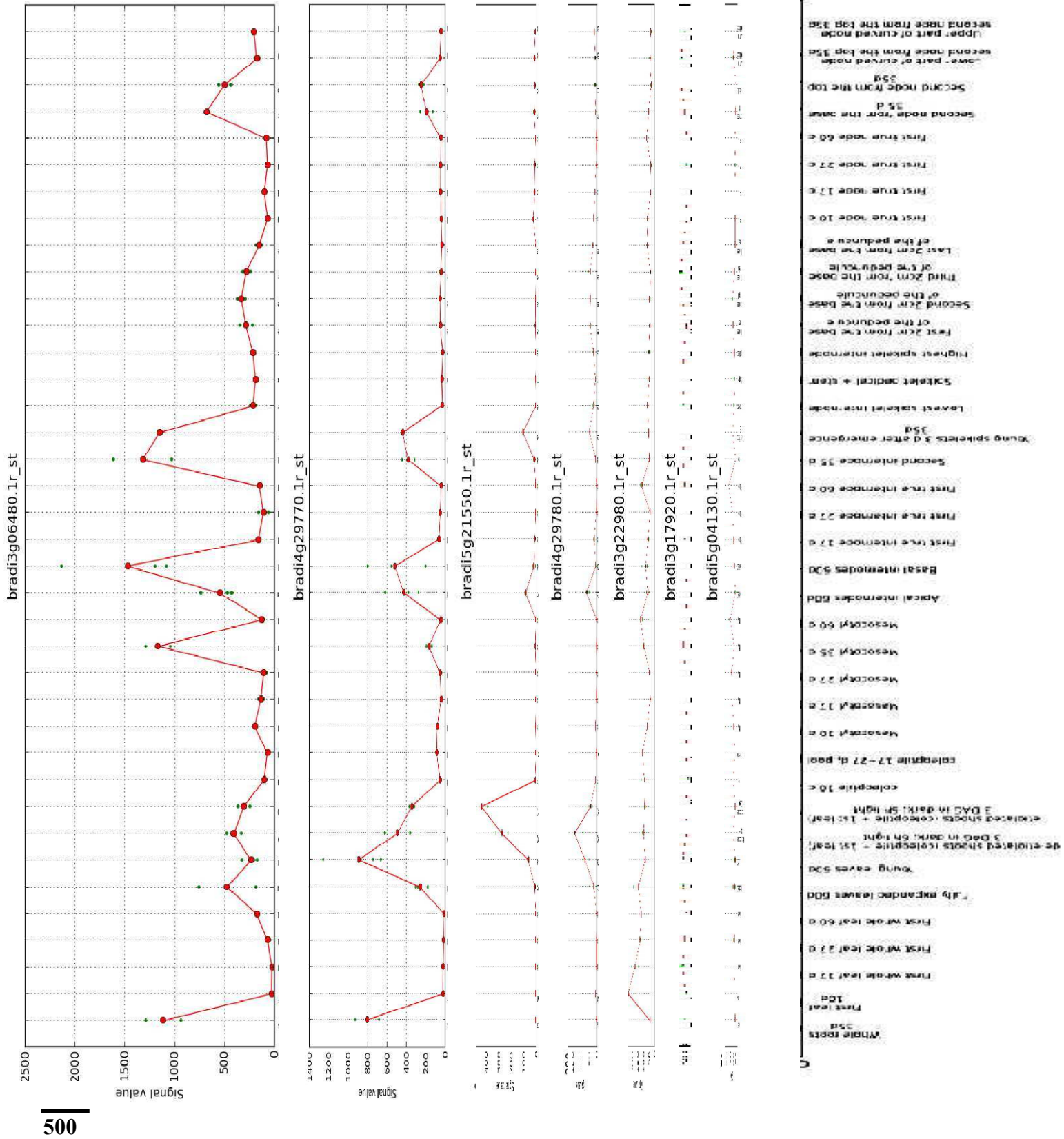
80 : flavonoïdes et phénylpropanoïdes

22 : isoprenoïdes et phénylpropanoïdes

87 : métabolisme des lipides ou des acides aminés

### 3. Profil d'expression par hybridation *in situ*

Afin de vérifier en détails la localisation de l'expression de *BdCAD1*, des *entre-nœuds* de Bd21-3 âgés de 9 semaines au stade de développement suivant l'émergence de l'épi, ont été observés après hybridation in-situ avec des sondes ciblant les ARN messagers de *BdCAD1* (Fig. 6). Les résultats des hybridations avec les sondes sens (Fig. 6b, d, f) montrent un marquage dû au bruit de fond. L'hybridation avec la sonde SamDC anti-sens (Fig. c) est le contrôle positif montrant que l'hybridation a eu lieu. Les mutants *cad* obtenus dans la collection de Versailles n'étant pas des mutants nuls, le contrôle négatif correspondant n'a pas



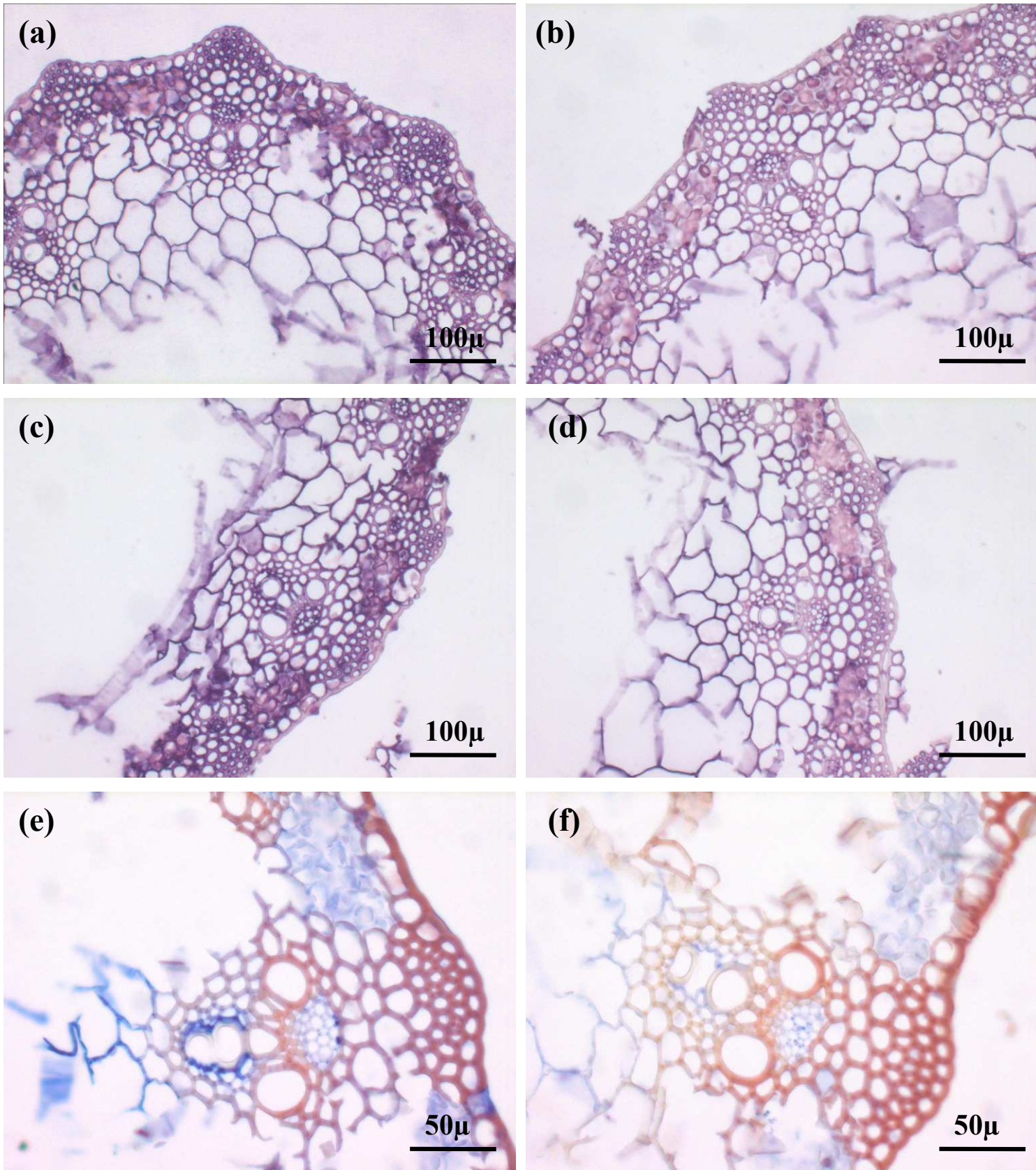
**Figure 5.** Profils d'expression des 7 CAD de *Brachypodium* en fonction des tissus. Les échelles ont été modifiées pour représenter la même valeur pour une même longueur. Source: Bradinet. Echelle rapportée à 500 unités arbitraires du signal d'expression.

pu être obtenu. L'hybridation avec la sonde CAD anti-sens, révélée à la Streptavidine Phosphatase Alcaline (SPA) (Fig. 6a) ou au NBT (Fig. 6e) est très similaire à celle obtenue avec SamDC. Ceci suggère, soit que BdcAD1 est un gène ubiquitaire dans l'*entre-nœud*, soit que la sonde utilisée n'est pas spécifique. La seconde hypothèse semble privilégiée par les observations faites sur les coupes transversale d'entre-nœuds des mutants qui suggéraient une spécificité de localisation de l'expression de BdcAD1.

Une construction contenant le promoteur de *BdcAD1* suivie du gène rapporteur *iudA* (GUS) a été générée et introduite dans Bd21-3 par transformation via *Agrobacterium tumefaciens*. Cette méthode a été employée par Sibout et al., (2003) pour montrer que AtCADC et AtCADD, les CAD impliquées dans la lignification chez *Arabidopsis thaliana*, sont exprimées dans les vaisseaux de xylème, le cambium interfasciculaire, et le cambium vasculaire. Les résultats à venir permettront peut être de déterminer plus précisément la localisation de l'expression de *BdcAD1* chez *Brachypodium distachyon*.

**Tableau 1.** Gènes *CAD* annotés chez *Brachypodium distachyon*. Correspondance entre leur nom, leur nom simplifié et les clusters de gènes auxquels ils appartiennent.

Nom simplifié	Nom du gène	Cluster	EST
BdCAD1	Bradi3g06480.1	94	3
BdCAD2	Bradi3g17920.1	87	0
BdCAD3	Bradi3g22980.4	132	5
BdCAD4	Bradi4g29770.1	61	10
BdCAD5	Bradi4g29780.1	80	0
BdCAD6	Bradi5g04130.1	Non renseigné	2
BdCAD7	Bradi5g21550.1	22	1



**Figure 6.** (a) (b) (c) (d) révélation à la SPA (e) (f) révélation au NBT de coupes de 8µ de la partie centrale d'entre-nœuds en cours de lignification de Bd21-3 en début de stade reproductif hybridées avec: (a) (e) sonde CAD antisens, (b) (f) sonde CAD sens, (c) sonde SamDC antisens, (d) sonde SamDC sens.



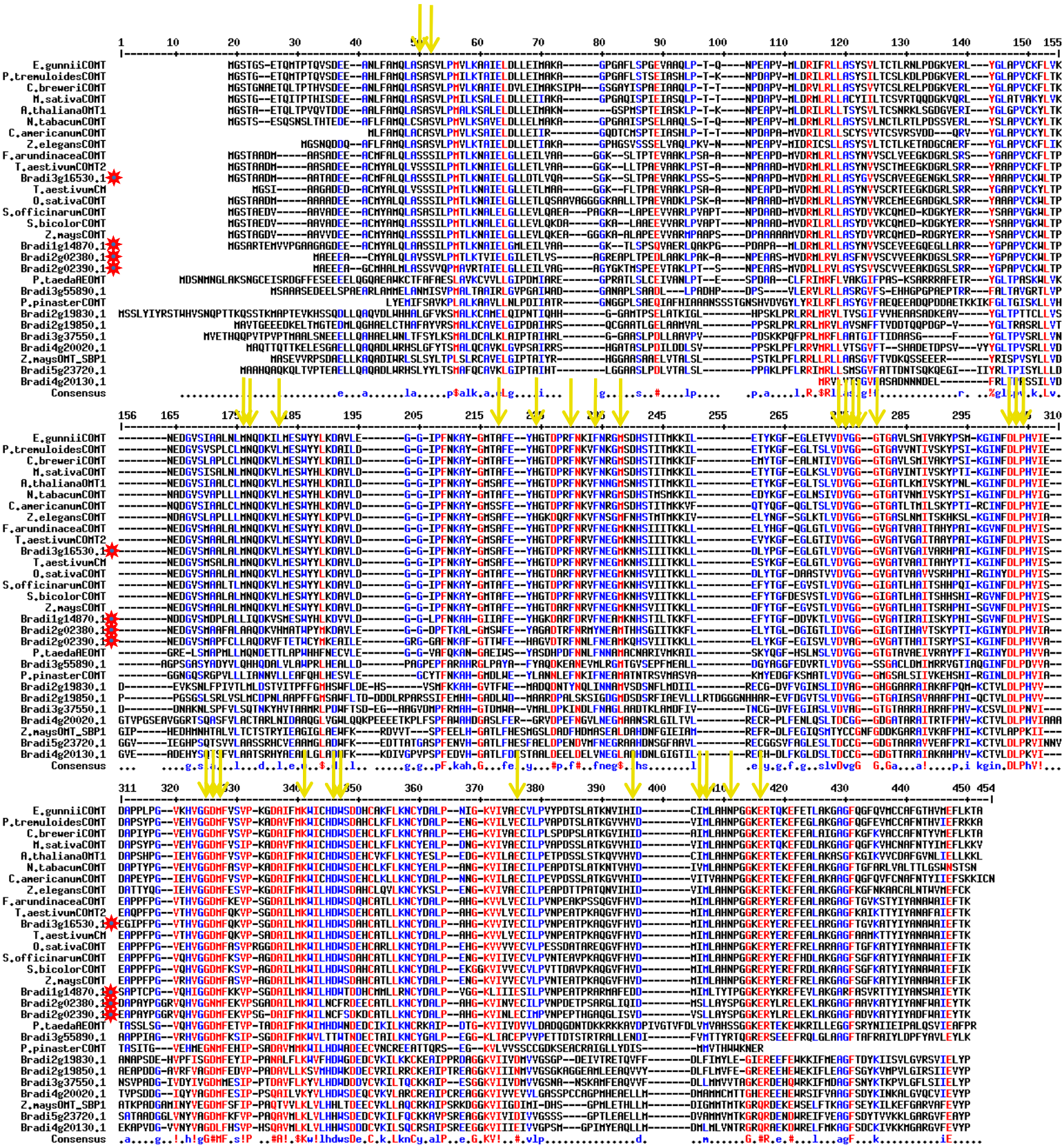


## Chapitre III :

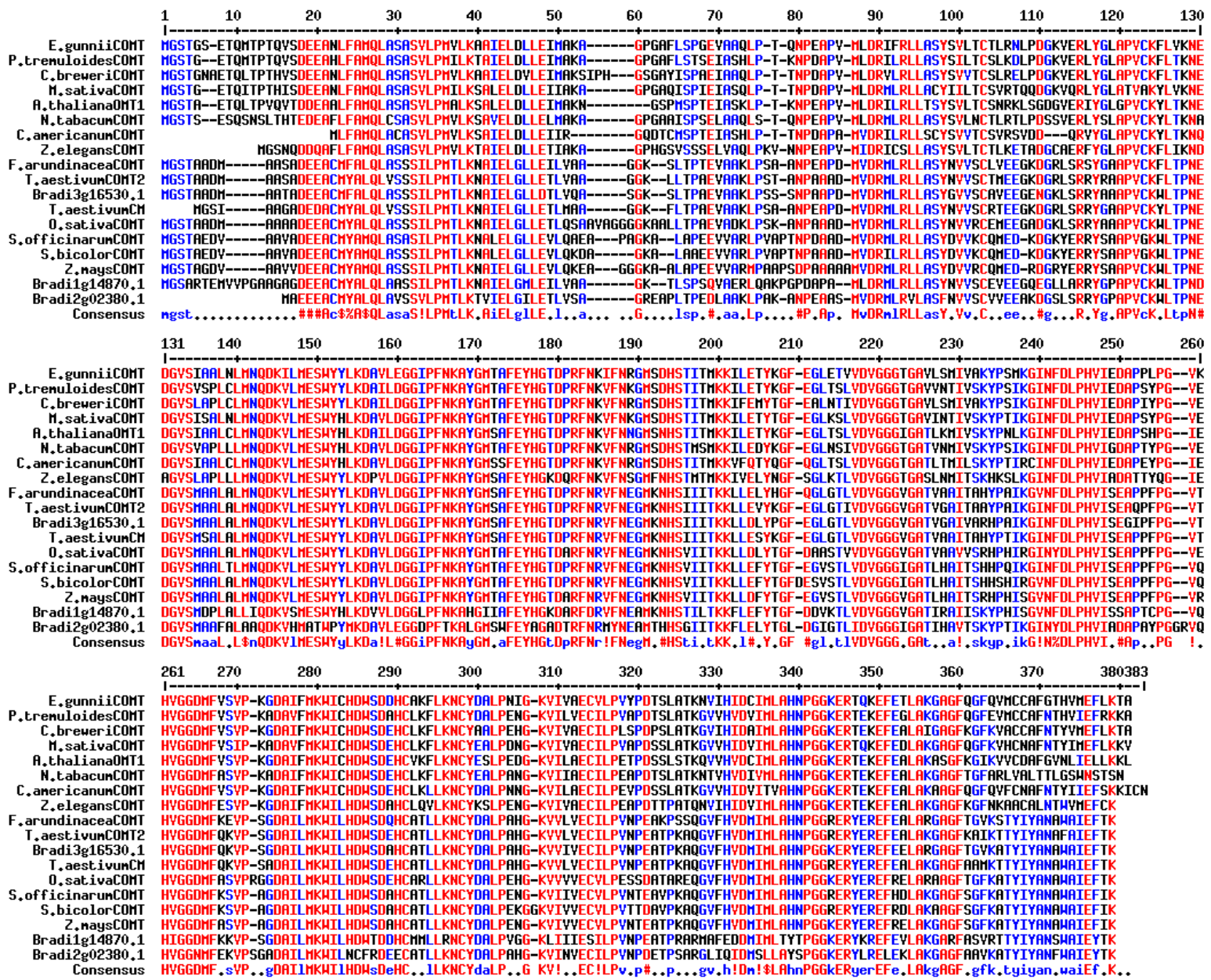
Identification de  
l'acide caféique *O*-méthyltransférase impliquée dans la  
voie de biosynthèse des monolignols

# Chapitre III : Identification de l'acide cafféique *O*- méthyltransférase impliquée dans la voie de biosynthèse des monolignols

Ce chapitre présente la détermination du gène candidat pour l'acide cafféique *O*-méthyltransférase impliquée dans la voie de biosynthèse des monolignols, la recherche de mutants dans ce gène et leur caractérisation.



**Figure 1.** Aligment de séquences protéiques des gènes de *Brachypodium* identifiés par BLAST avec des COMEY identifiées chez diverses espèces végétales. Les flèches jaunes indiquent les résidus essentiels à l'activité de la protéine et les étoiles rouges à droite des noms des protéines alignées indiquent les candidats COMEY de *Brachypodium*. En rouge: résidus à fort consensus; en bleu: résidus à consensus moyen, en noir les résidus neutres.



**Figure 2.** Alignment de séquences protéiques des candidats COMT de *Brachypodium* avec des COMT de diverses espèces végétales. La présence d'un plus grand nombre de résidus à fort consensus (en rouge) comparé à l'alignement précédent confirme le choix des candidats. En bleu: résidus à consensus moyen, en noir les résidus neutres

L'acide cafféique *O*-méthyltransférase catalyse la méthylation du Carbone 5 du 5-hydroxyconiféraldéhyde en sinapaldéhyde, dirigeant la voie de biosynthèse des monolignols vers la formation des unités S des lignines. Les plantes déficientes pour ce gène présentent des modifications importantes de leur teneur et composition en lignines et acides hydroxycinnamyliques, mais peu de défauts de croissance, ce qui en fait de bons candidats pour l'amélioration de la biomasse dédiée à la production d'éthanol-carburant à partir des parois cellulosiques.

### 1. Détermination du gène candidat par alignements de séquences

Parmi les gènes *COMT* dont l'implication dans la voie de biosynthèse des monolignols a été montrée, la traduction en séquence protéique de la *COMT* d'*Arabidopsis thaliana* (*AtOMT1* Locus: At5g54160) a été choisie pour servir d'entrée au BLAST sur le génome 4X, puis 8X de *Brachypodium distachyon*. Parmi les gènes de *Brachypodium* issus de cette recherche, les dix possédant les scores les plus élevés et les e-value les plus faibles sont les suivants :

	Score	P(N)	N
Bradi3g16530.1	1170	6.7e-120	1
Bradi1g14870.1	1048	5.7e-107	1
Bradi2g02380.1	953	6.6e-97	1
Bradi2g02390.1	947	2.9e-96	1
Bradi3g55890.1	572	1.6e-56	1
Bradi2g19830.1	466	2.7e-45	1
Bradi3g37550.1	404	1.0e-38	1
Bradi4g20020.1	355	1.6e-33	1
Bradi2g19850.1	333	3.3e-31	1
Bradi4g20130.1	323	3.8e-30	1

Parmi ceux-ci, seules les valeurs associées aux quatre premiers gènes sont satisfaisantes pour estimer que la protéine pour laquelle ils codent possède une fonction identique ou similaire à

**Tableau 1.** Correspondance entre les appellations des gènes *COMT* annotés chez *Brachypodium distachyon* et résultats de la recherche sur les banques d'EST à partir des séquences des trois gènes *BdCOMT*.

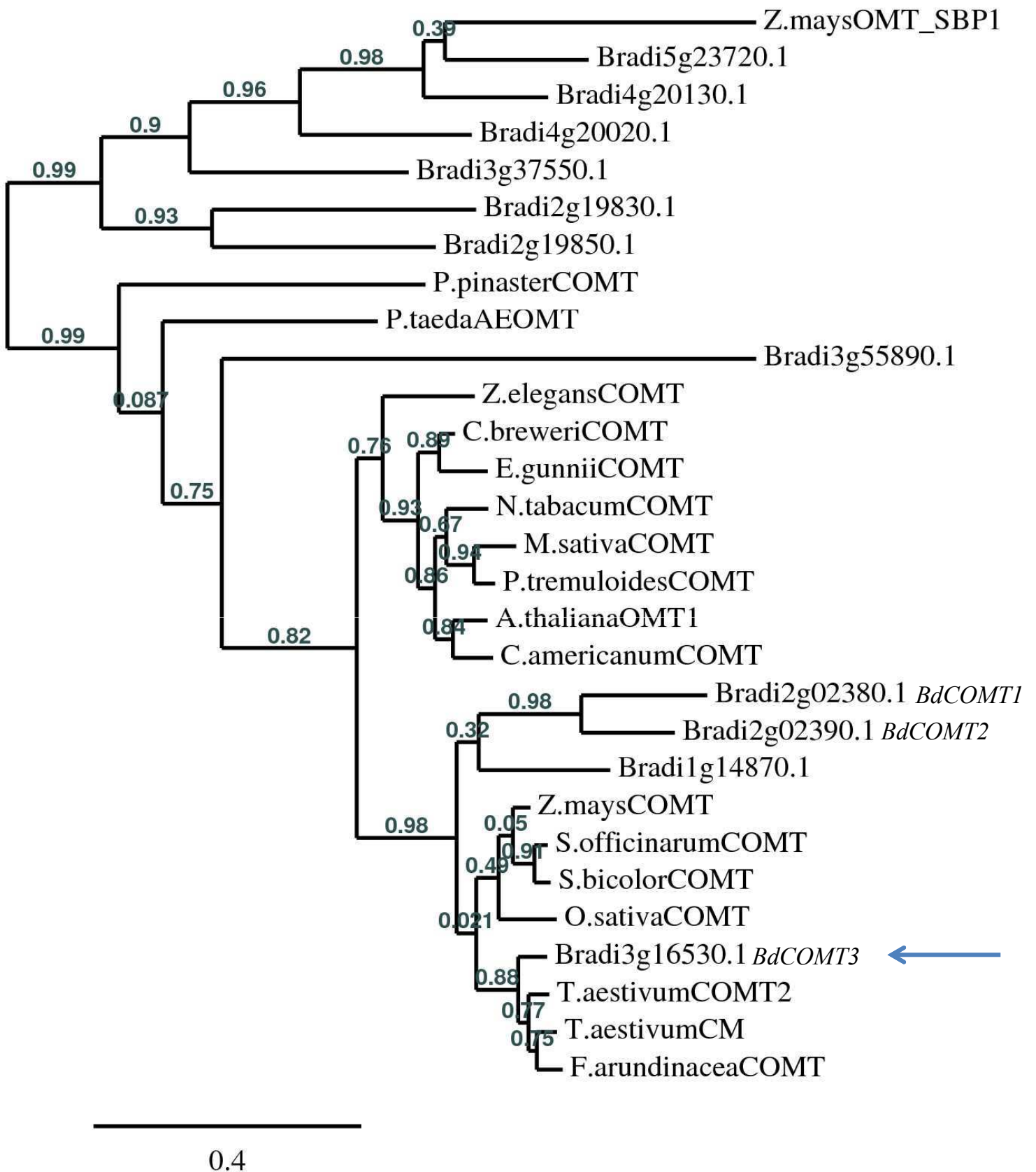
Nom d'après leur position sur le génome	Nomenclature simplifiée	EST
<a href="#">Bradi2g02380.1</a>	<i>BdCOMT1</i>	0
<a href="#">Bradi2g02390.1</a>	<i>BdCOMT2</i>	0
<a href="#">Bradi3g16530.1</a>	<i>BdCOMT3</i>	24

AtCOMT1 et seront privilégiés dans la suite de l'analyse. Il est à noter que l'utilisation d'une autre séquence pour effectuer ce BLAST (COMT de maïs ou de ray-grass par exemple) donne une liste identique.

Les dix protéines ont été alignées à un certain nombre d'autres protéines COMT de diverses espèces (Fig. 1). La longueur des séquences est le premier élément observé. Si l'une d'entre elle était particulièrement plus courte ou plus longue que la moyenne pour cette famille de gènes, sa fonction serait remise en question. Dans le cas présent, ce critère n'est pas déterminant. Sur cet alignement, les acides aminés particulièrement importants pour la fonction de la protéine et généralement très conservés ont été identifiés (Zubieta, Kota et al. 2002) et représentés par les flèches jaunes. En raison de différences trop marquées pour tous ces sites (acides aminés non conservés et sans propriétés physiques similaires) cet alignement a permis d'éliminer les protéines autres que celles codées par les quatre premiers gènes : à ce niveau, nos candidats étaient donc Bradi3g16530.1, Bradi1g14870.1, Bradi2g02380.1 et Bradi2g02390.1. Cependant, seul Bradi3g16530.1 présente une conservation exacte de ces sites, les trois autres montrant des différences importantes au niveau des acides aminés de positionnement et de fixation du substrat sur le site actif, et pour Bradi2g02390.1 également sur le site de fixation au cofacteur (Fig. 2).

Depuis cette recherche, le MIPS (German Research Center for environmental Health) a fourni une annotation pour les COMT de *Brachypodium* en ne prenant en compte que *Bradi2g02380.1*, *Bradi2g02390.1* et *Bradi3g16530.1* qui seront respectivement appelés *BdCOMT1*, *BdCOMT2* et *BdCOMT3* selon leur position sur le génome de *Brachypodium* (Tableau 1). Le gène *Bradi1g14870.1* n'est pas annoté en tant que *COMT* sur ce site malgré sa similitude de séquence avec les *COMT* annotées.





**Figure 3.** Arbre phylogénétique construit à partir des séquences protéiques de COMT incluant celles de *Brachypodium distachyon*. Les chiffres gris indiquent la valeur « bootstrap » représentant la probabilité d'existence de chaque branche du phylogramme lors des répétitions des calculs de construction. La longueur des branches est proportionnelle à la distance évolutive entre les nœuds. La flèche bleue indique le gène de *Brachypodium* choisi comme candidat.

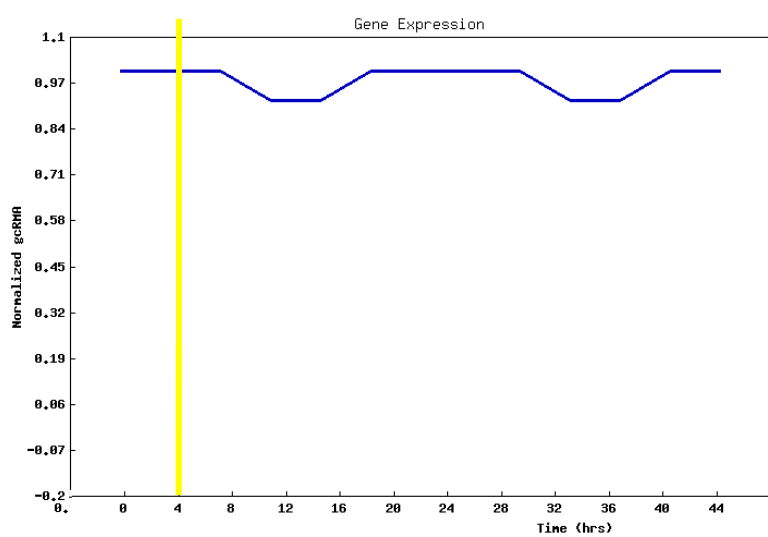
## 2. Etude phylogénétique

L'intégralité des gènes utilisés pour le premier alignement a ensuite été soumise à une analyse phylogénétique sur phylogeny.fr (Fig. 3). Sur l'arbre phylogénétique, deux groupes se distinguent : les *COMT* de dicotylédones d'une part, et celles de graminées d'autre part. Nos quatre candidats se trouvent dans le second groupe, cependant, Bradi3g16530.1 est intégré dans un sous-groupe contenant les *COMT* de blé et de fétuque, tandis que les trois autres forment un sous-groupe distinct sous la branche des graminées (*BdCOMT1*, *BdCOMT3* et *Bradi1g14870.1*). Ceci confirme notre hypothèse selon laquelle *Bradi3g16530.1* serait le candidat le plus probable pour la voie de biosynthèse des monolignols.

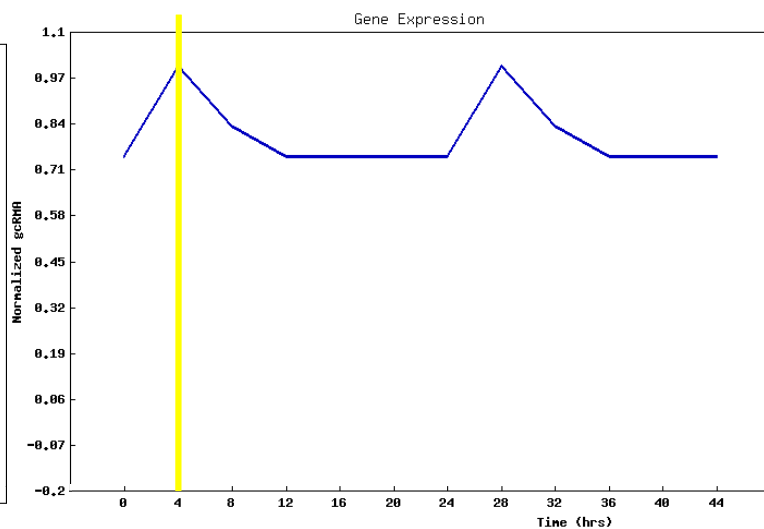
Les EST (pour Expressed Sequence Tag), représentant une indication du niveau d'expression d'un gène, ont été recherchées pour les trois *BdCOMT* (Tableau 1). Il s'agit de la recherche de séquences cibles sur les banques d'EST qui réunissent les données quantitatives de séquençage de courts fragments provenant de banques d'ADNc. L'existence d'EST correspondant à un gène montre que celui-ci est exprimé (Nagaraj, Gasser et al. 2007). Il y a actuellement plus de 218000 EST disponibles sur le génome de *Brachypodium* ([http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html)).

## 3. Expression en fonction du cycle circadien

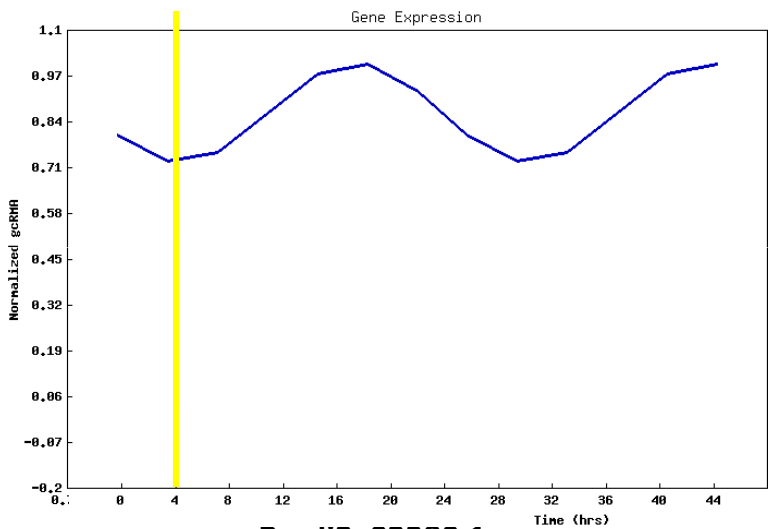
Les données obtenues sur Diurnal (Fig. 4) ne permettent pas de modifier les conclusions précédentes. On observe que *BdCAD3* (Fig. 4a) et *BdCAD2* (Fig. 4d) présentent un maximum d'expression 4 heures après exposition à la lumière. Cependant, aucun ne présente un pic d'expression comme pour *Arabidopsis* (Fig. 4b).



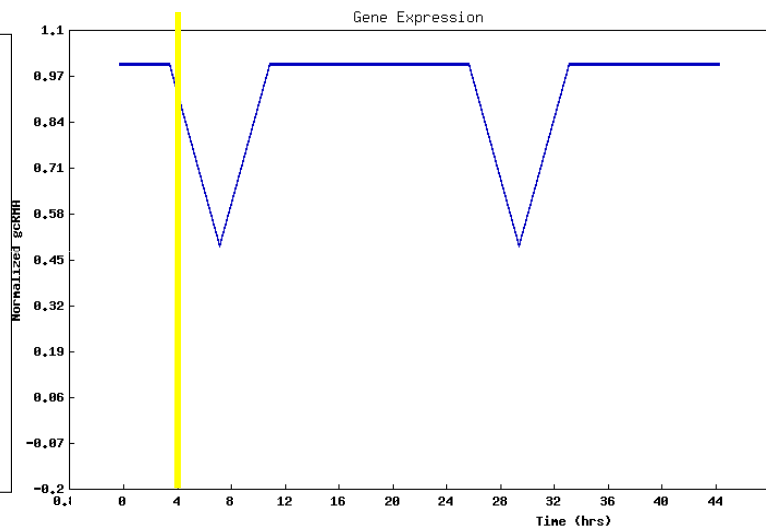
**Bradi5g16530.1**



**At5g54160**



**Bradi2g02380.1**



**Bradi2g02390.1**

**Figure 4.** Expression en fonction du cycle circadien des gènes *BdCOMT3* (a), *AtOMT1* (b), *BdCOMT1* (c) et *BdCOMT2* (d) dans les conditions de culture les plus proches de celles de Versailles. Le temps en heures est indiqué en abscisse, et une unité arbitraire normalisée est indiquée en ordonnée. Les barres jaunes indiquent quatre heures après le début de l'exposition à la lumière.

## 4. Profils d'expression en fonction des tissus

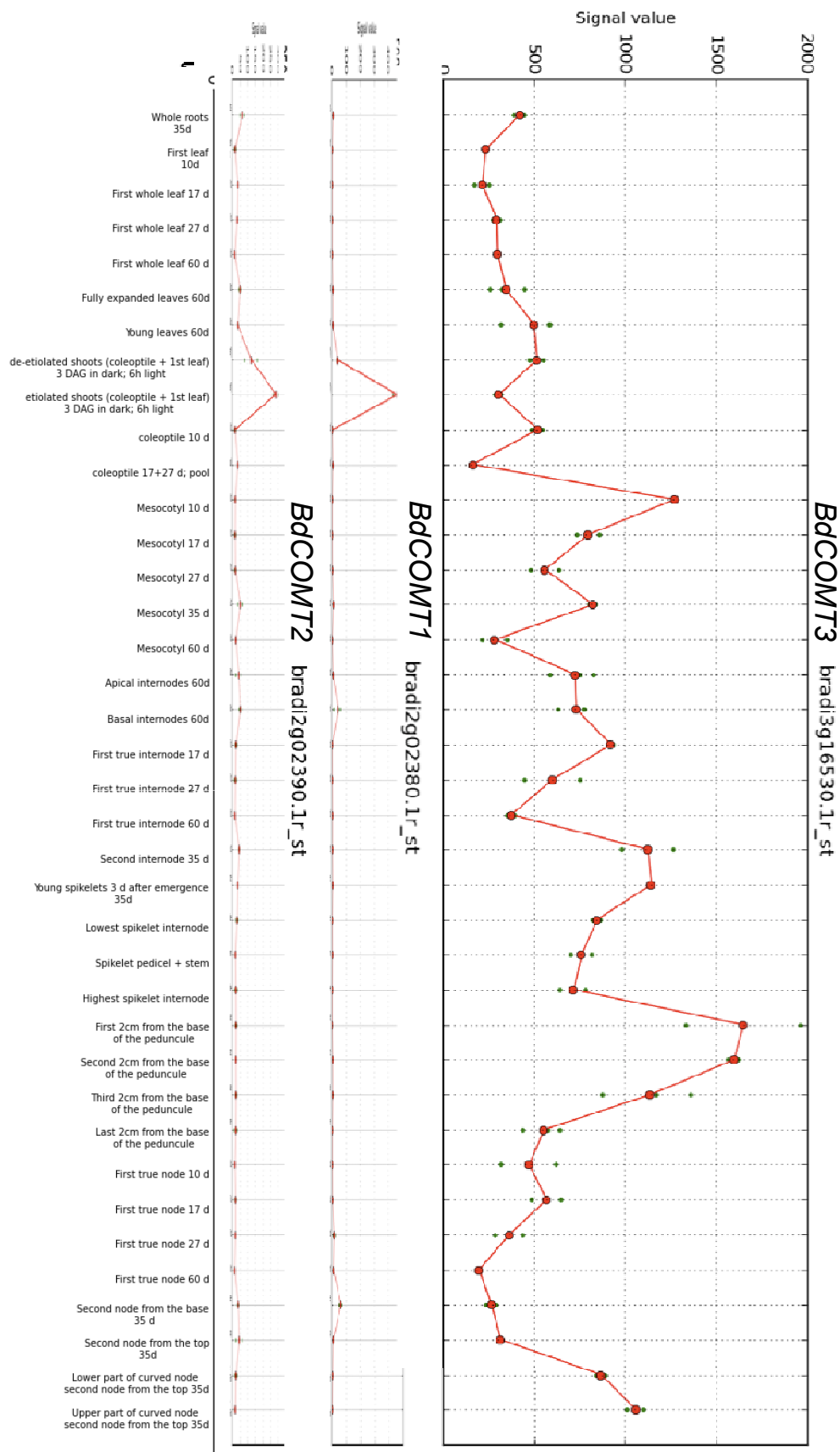
### 4.1. Bradinet

Les données obtenues sur Bradinet (Fig. 5) indiquent que *BdCOMT3* est la *COMT* la plus exprimée. Les tissus présentant les plus fortes expressions de *BdCOMT3* sont le mésocotyle, les entre-nœuds, le pédoncule (particulièrement sa partie basse, celle en cours de lignification), et le nœud. On peut observer que pour la plupart des tissus, le niveau d'expression le plus fort est à 3 jours après émergence de l'épi. Le gène Bradi1g14870.1 présente un niveau d'expression extrêmement faible et est donc exclu de la liste des candidats *COMT*. Par ailleurs, les données de coexpression des trois *COMT* annotées indiquent que *BdCOMT3* est la plus coexprimée avec d'autres gènes de la voie de biosynthèse des lignines.

### 4.2. Etudes d'expression par RT-PCR

Les ARN de tiges (entre-nœuds, nœuds et gaine foliaire) ont été extraits aux quatre stades décrits dans le chapitre précédent (à cinq, sept, neuf et douze semaines de culture), rétrotranscrits et amplifiés par RT-PCR semi-quantitative avec des oligonucléotides spécifiques du gène *BdCOMT3* et du gène constitutif *SamDC* (Hong, Seo et al. 2008) (Fig. 6a). *BdCOMT3* est fortement exprimé à tous les stades, avec un niveau très légèrement supérieur juste après émergence de l'épi (stade R), puis légèrement plus faible avant la sénescence (stade S). Ces résultats sont cohérents avec l'expression d'un gène impliqué dans la voie de biosynthèse des monolignols.

Les ARN de différents organes du brin maître de trois plantes Bd21-3 ont été mis en commun et rétrotranscrits avec des oligonucléotides spécifiques du gène *BdCOMT3* et du gène constitutif *SamDC* (Fig. 6b). *BdCOMT3* est fortement exprimé dans chacun des organes : entre-nœuds jeunes en cours d'élongation (correspondant au premier centimètre à partir du nœud des trois entre-nœuds les plus hauts), entre-nœuds matures (*entre-nœuds* basals entiers), nœuds,



**Figure 5.** Niveaux d'expression des trois *COMT* annotées de *Brachypodium* dans des organes différents. Données provenant de Bradinet. Afin de faciliter les comparaisons, les trois graphes sont à la même échelle en ordonné (échelle indiquant 500 unités d'expression). Les différents tissus sont indiqués en abscisse.

gaines foliaires, pédoncule et épis, à l'exception des feuilles où il est beaucoup plus faiblement exprimé. Ceci a été vérifié par RT-PCR quantitative sur un pool de quatre maîtres brins de Bd21-3 au stade R comparé aux feuilles des mêmes brins où *BdCOMT3* est en moyenne quatre fois moins exprimé. Les tiges étant les organes les plus lignifiés de la partie aérienne des graminées, et les feuilles étant les organes les moins lignifiés, ces résultats sont cohérents avec l'implication de ce gène dans la lignification.

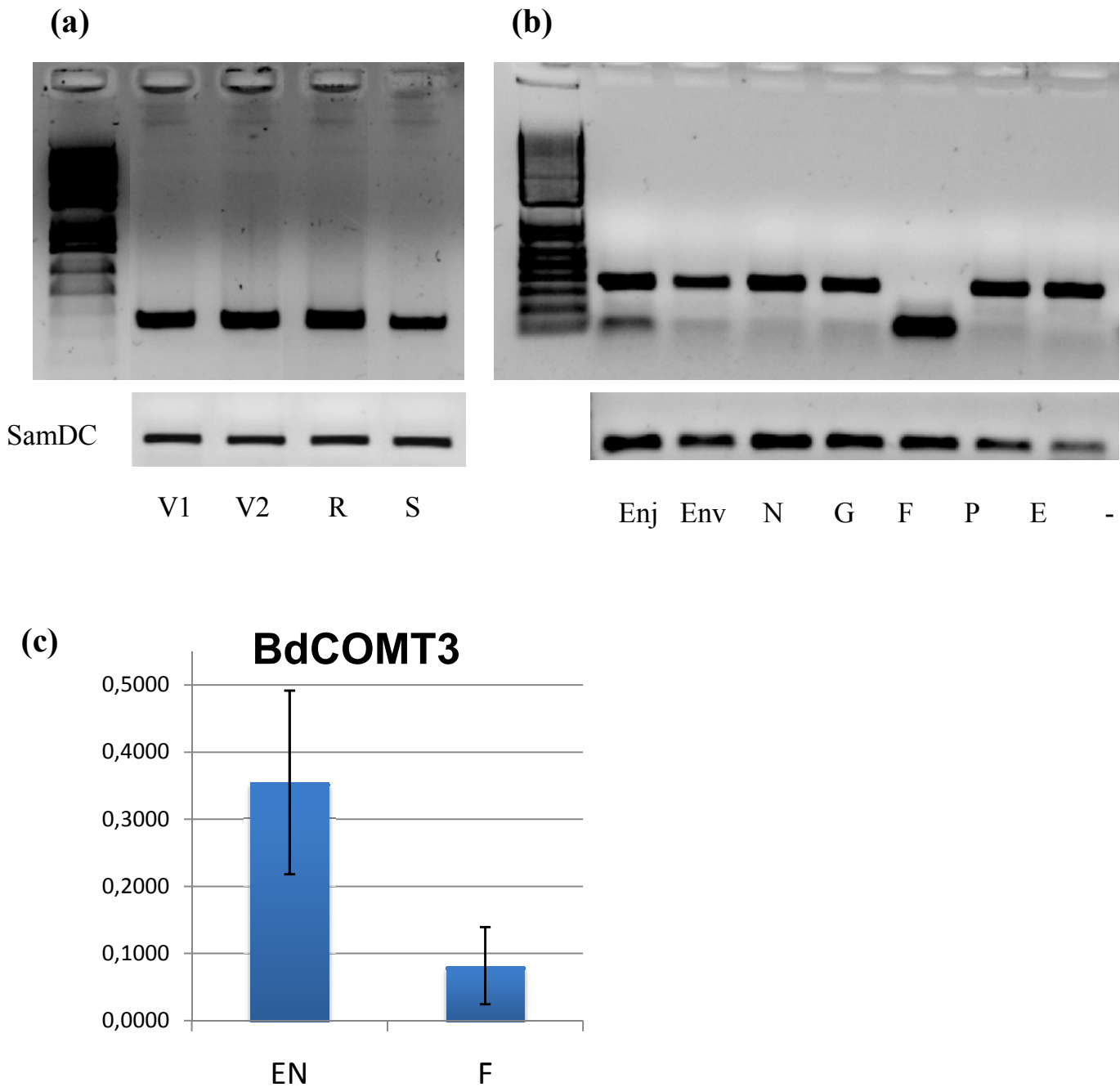
## 5. Complémentation fonctionnelle

Afin de déterminer si la protéine codée par le gène *BdCOMT3* possède une activité acide cafféique *O*-méthyltransférase, il a été cloné sous contrôle du promoteur constitutif 35S du CaMV puis introduit dans la lignée *DML6* d'*Arabidopsis thaliana*, lignée mutante pour le gène *AtOMT1* (Goujon, Sibout et al. 2003).

### 5.1. Analyses des lignines

Après sélection et culture de la première génération (T1) issue des plantes transformées, leur lignine a été analysée par thioacidolyse (Tableau 2).

Pour les tiges du témoin WS, le rapport S/G, correspondant aux proportions relatives des deux principales unités des lignines lorsqu'elles sont uniquement liées par des liaisons  $\beta$ -O-4, présente une valeur de 0,45, valeur cohérente pour ce type d'échantillon. Chez le mutant *DML6*, ce rapport chute à une valeur caractéristique de 0,01 (Goujon, Sibout et al. 2003). Cet effondrement des monomères S de thioacidolyse confirme que la déficience en COMT bloque en grande partie la biosynthèse des unités S. Parallèlement à cet effondrement en unités S, on observe un enrichissement en unités G et surtout l'apparition de monomères de thioacidolyse issus des unités 5-hydroxyguaïacyles (5-OH G, Tableau 2). Dans le cas des échantillons T1, on observe que le rapport S/G est effondré pour un seul échantillon, les autres présentant des



**Fig 6.** (a) RT-PCR semi quantitative sur les tiges de Bd21-3 quatre stades décrits au chapitre II: fin du stade végétatif à 5 semaines (V1), avant émergence de l'épi à 7 semaines (V2), après émergence de l'épi à 9 semaines (R) et avant sénescence à 12 semaines (S). Les trois premiers stades apparaissent similaires, avec peut être une expression un peu plus intense à 9 semaines. Une seule de 3 répétitions biologiques effectuées est montrée. (b) RT-PCR au stade 7 semaines sur différents organes du maître brin de trois plantes. Enj: jeunes entre-nœuds (en cours de lignification); Env: entre-nœuds lignifiés; N: nœuds; G: gaine foliaire; F: feuilles; P: pédoncule; E: épi. (c) RT-PCR quantitative montrant les niveaux d'amplification de BdCOMT3 rapportés à ceux du gène de ménage SamDC dans les entre-nœuds (EN) et les feuilles (F) de quatre répétitions biologiques de Bd21- 3 après émergence de l'épi. (a) (b) L'amplification du gène constitutif SamDC permet d'indiquer la quantité de matrice présente dans chaque puits. (c) Les données ont été normalisées par rapport au gène constitutif SamDC.

valeurs de S/G intermédiaires entre WS et DML6. Mise à part l'une d'entre elles (*comt16*), toutes les lignées complémentées présentent des valeurs intermédiaires entre celles du mutant et celles du sauvage : leur rapport S/G avoisine 0,20. Ces données suggèrent que la complémentation fonctionnelle a bien eu lieu mais seulement de façon partielle. Par la suite, des lignées homozygotes T3 avec un seul site d'insertion de l'ADN-T ont été sélectionnées, cultivées en plus grande quantité et leurs tiges mûres ont été analysées en termes de quantité de lignines (évaluée par le % de Lignine Klason dans les tiges mûres débarrassées des extractibles) et en terme de structure des lignines (évaluée par thioacidolyse). Ces analyses plus approfondies ont pu être réalisées en doubles analytiques (Tableau 3). Les différents échantillons (WS, DML6 et les lignées complémentées homozygotes) présentent des taux de lignines très similaires, ce qui confirme que la déficience en activité COMT ne modifie pas la teneur en lignines des tiges d'arabette (Goujon, Sibout et al. 2003). Les données du Tableau 3 révèlent en outre que ni le rapport S/G, ni le rendement en monomères issus des unités 5-OH G ne permettent de discriminer les lignées complémentées du mutant DML6 (Tableau 3). Ce résultat suggère fortement l'hypothèse du silencing dû à la force du promoteur et à l'insertion d'un gène exogène très homologue du gène endogène. En outre, pour toutes les lignées déficientes en activité COMT (DML6 et mutants complémentés), on constate que le rendement de thioacidolyse est systématiquement plus faible que chez le témoin WS, lorsque ce rendement est calculé en micromoles par gramme de lignine Klason. Ce résultat confirme que les lignines de ces plantes sont enrichies en liaisons condensées (résistantes à la thioacidolyse).

Malgré ce phénomène inattendu de silencing, l'analyse des plantes T1 a montré la complémentation partielle du mutant DML6 par le gène *BdCOMT3*. Ce dernier code bien pour une protéine possédant une activité *O*-méthyltransférase impliquée dans la lignification.



**Tableau 2.** Etude par thioacidolyse sur matériel non extrait des lignines des tiges matures de la première génération (T1) de lignées DML6 complétées (*comtcp*). Une seule plante de chaque lignée complétée a été analysée, et quatre plantes pour le témoin sauvage WS et pour la lignée mutante DML6 ont été réunies et analysées en tant qu'échantillon. Le rendement en monomères *p*-hydroxyphényles (H), guaiacyles (G), syringyles (S) et 5-hydroxyguaïacyles (5-OH G) est exprimé en µmoles par gramme de matériel végétal sec analysé.

Echantillon	Rendement en µmole/g de matériel végétal					S/G
	H	G	S	total	5OH_G	
WS	0,8	81	36	118	0,4	0,45
DML6	0,5	85	1	86	3,4	0,01
WS+Bd4	0,6	92	19	112	1,7	0,21
comtCp 10 (DML6+Bd4)	0,7	98	19	117	1,9	0,19
comtCp 12 (DML6+Bd4)	0,5	73	8	81	1,4	0,10
comtCp 15 (DML6+Bd4)	0,7	92	23	116	1,5	0,25
comtCp 16 (DML6+Bd4)	0,6	88	1	90	3,0	0,01
comtCp 24 (DML6+Bd4)	0,8	100	19	120	1,8	0,19

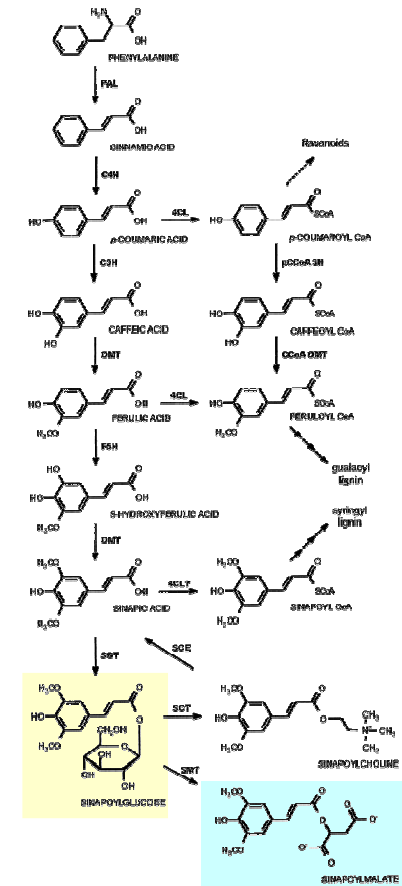
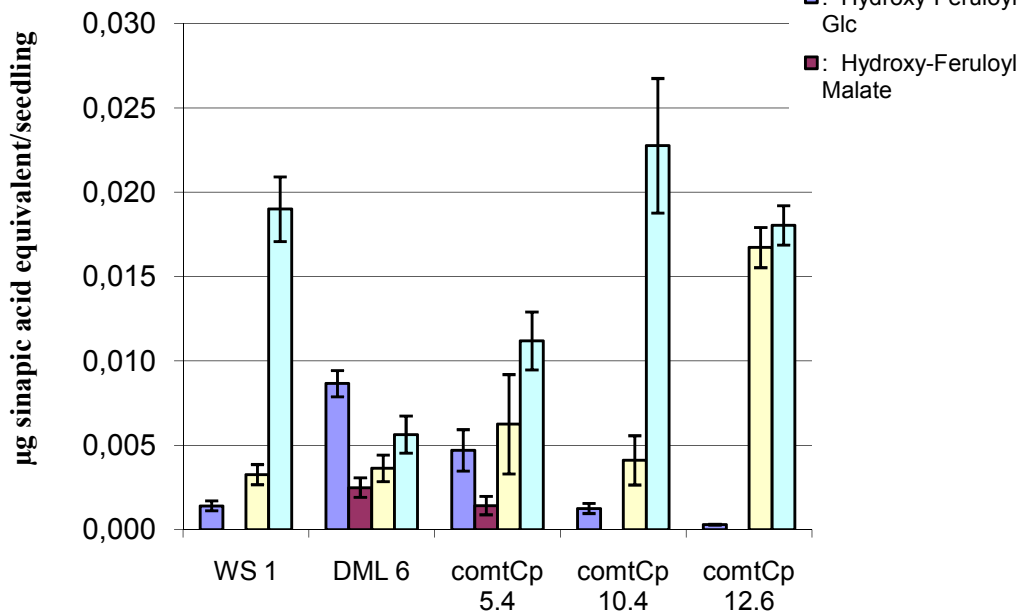
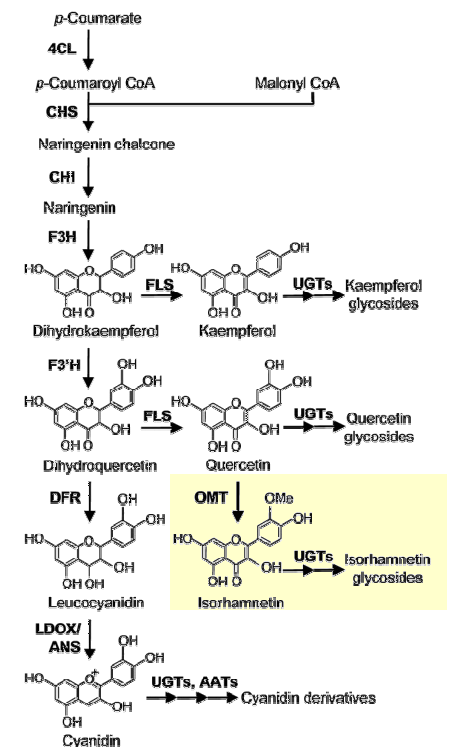
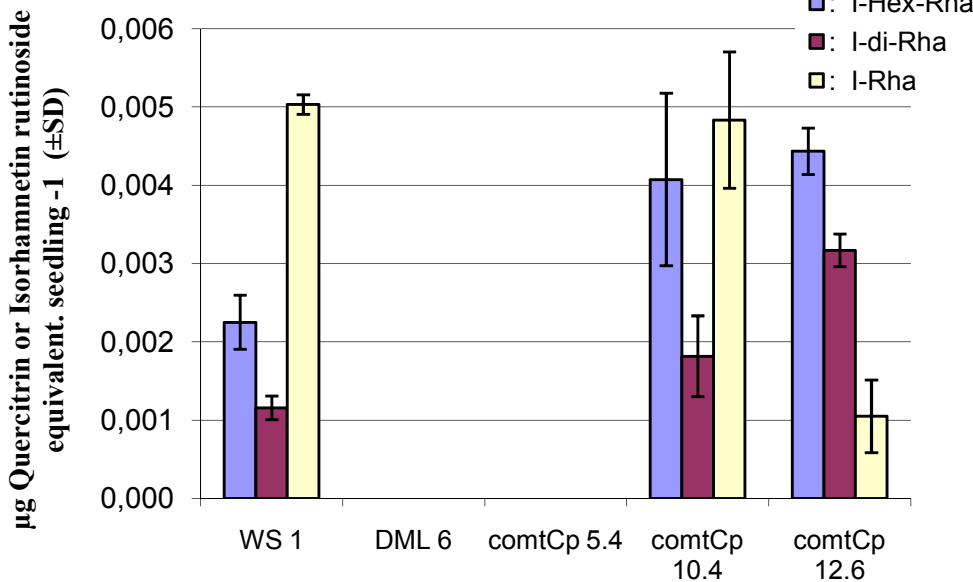
**Tableau 3.** Etude des lignines des tiges matures de la troisième génération (T3) de lignées DML6 complétées par *BdCOMT3* (*comtcp*). Les résultats correspondent aux valeurs moyennes et aux erreurs moyennes entre doubles biologiques (2 lots de 4 plantes par génotype). Le taux de lignine Klason (%LK) est exprimé en % pondéral du résidu pariétal (échantillon débarrassé de ses composés solubles). Le rendement total en monomères de thioacidolyse est exprimé en µmoles par gramme de LK.

Echantillon	%LK	Rendement en µmole/g LK					S/G
		H	G	S	5-OHG	total	
WS	19.80 ± 0.13	5.7 ± 0.3	1018 ± 4	371 ± 6	0.5 ± 0,0	1396 ± 9	0.36 ± 0,00
DML6 <i>comt1</i>	18.16 ± 0.28	4.7 ± 0.3	1062 ± 86	5 ± 1	4.8 ± 0.3	1077 ± 88	0.01 ± 0,00
WS + <i>BdCOMT3</i>	18.48 ± 0.67	4.3 ± 0.1	1126 ± 1	15 ± 0	6.8 ± 0.3	1152 ± 1	0.01 ± 0,00
comtCp 10-4	19.22 ± 0.11	5.0 ± 0.5	1003 ± 33	157 ± 16	3.5 ± 1,0	1168 ± 19	0.16 ± 0.02
comtCp 12-6	19.62 ± 0.04	4.6 ± 0.5	1029 ± 37	84 ± 9	5.0 ± 0.1	1123 ± 28	0.08 ± 0.01
comtCp 15-2	18.16 ± 0.06	4.9 ± 0.2	1041 ± 7	7 ± 2	7.4 ± 1,0	1061 ± 10	0.01 ± 0,00

## 5.2. Composés phénoliques solubles

Les enzymes de la voie de biosynthèse des monolignols peuvent également jouer un rôle dans la biosynthèse de composés phénoliques solubles, tels que les dérivés de l'acide sinapique (sinapoyl malate et sinapoyl glucose) caractéristiques des brassicacées, et certains flavonoïdes méthylés, tels que les glycosides d'isorhamnétine. Ainsi, des travaux antérieurs ont révélé que la déficience en activité COMT provoque une diminution du taux de sinapoyl malate dans les tiges adultes ou dans les plantules du mutant *DML6*, diminution accompagnée de l'apparition de 5-OH-féruloyl malate (5-OH-Fe-M) et de 5-OH-féruloyl glucose (5-OH-Fe-Glc) (Goujon, Sibout et al. 2003).

La présence de ces composés a été étudiée chez des plantules (T3) âgées de 10 jours (culture in vitro sur boîte de Pétri). Les résultats montrent qu'il y a accumulation de 5-OH-Fe-M et de 5-OH-Fe-Glc ainsi qu'une diminution de quantité de sinapoyl malate chez le mutant *DML6* par rapport au sauvage, mais pas de différence significative au niveau du sinapoyl glucose. Dans la lignée sauvage WS surexprimant *BdCOMT3*, les trois premiers composés s'accumulent à des valeurs intermédiaires entre celles du sauvage et du mutant, suggérant encore le phénomène de silencing sur cette lignée. Par contre, chez les lignées *DML6* complémentées, on observe une diminution de la présence de 5-OH-Fe-Glc, la disparition du 5-OH-Fe-M et une restauration du taux de sinapoyl malate aux valeurs obtenues pour la lignée sauvage (Fig. 7a). Il y a donc bien eu complémentation au niveau des voies impliquant ces composés. Concernant les glucosides d'isorhamnétines, ils sont totalement absents dans la lignée *DML6*, ainsi que dans la lignée WS + *BdCOMT3*, montrant que le silencing ayant lieu dans cette lignée diminue son expression à un point l'empêchant d'accomplir sa fonction au niveau de ces composés (Fig.7b). Dans les deux lignées mutantes complémentées, on observe par contre que *BdCOMT3* a efficacement converti les glycosides de quercétine en glycosides d'isorhamnétine correspondants.

**(a)****(b)**

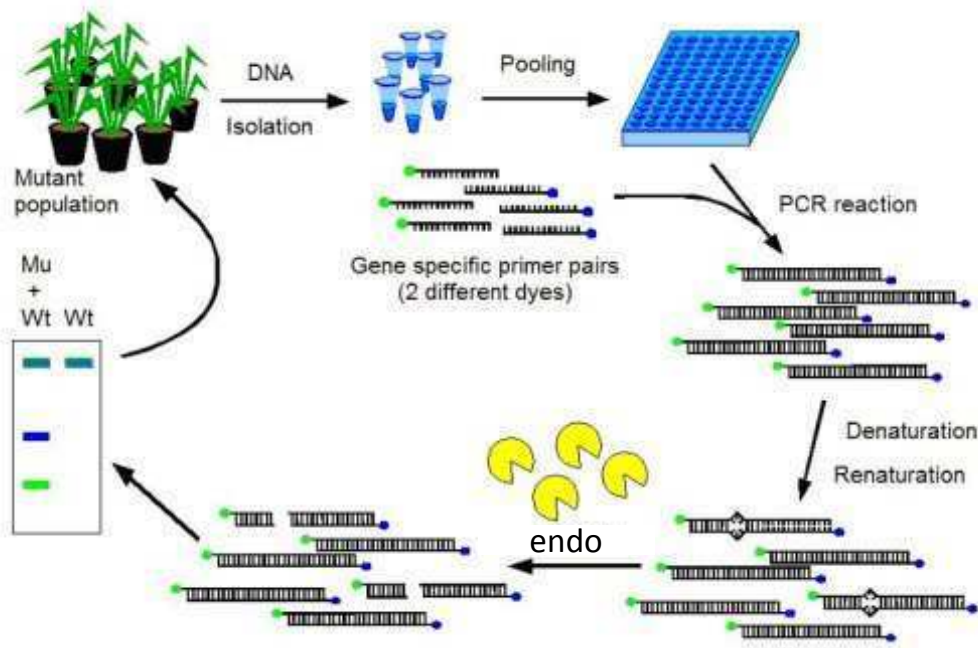
**Figure 7.** Analyses de phénols solubles dans des plantules T3 de 10 jours cultivées sur boîte de Pétri des lignées d'*Arabidopsis thaliana* sauvage (WS1), mutante (DML6) et ces lignées surexprimant le gène *BdCOMT3* (comtCp). (a) Teneurs en hydroxy-féruoyl glucose, hydroxy-féruoyl malate, sinapine (sinapoyl Glc) et en sinapoyl malate et voie de biosynthèse menant à leur formation (Ruegger et al. 1999). Les résultats correspondent aux moyennes et écarts-types entre trois répétitions et sont exprimés en équivalent d'acide sinapique (en µg) par plantule (récoltée fraîche). (b) Teneurs en Isorhamnétine et en glycosides d'isorhamnétines et la voie de biosynthèse menant à leur formation (Yonekura-Sakakibara et al. 2008). Les résultats correspondent aux moyennes et écarts-types entre trois répétitions et sont exprimés en équivalent de quercétine (en µg) par plantule (récoltée fraîche).

Les résultats de la partie précédente suggéraient que BdCOMT3 avait l'activité COMT nécessaire à la voie de biosynthèse des monolignols. Ici, il est montré que cette COMT agit sur les mêmes substrats qu'AtOMT1, renforçant la position de *BdCOMT3* en tant que gène d'intérêt pour la lignification.

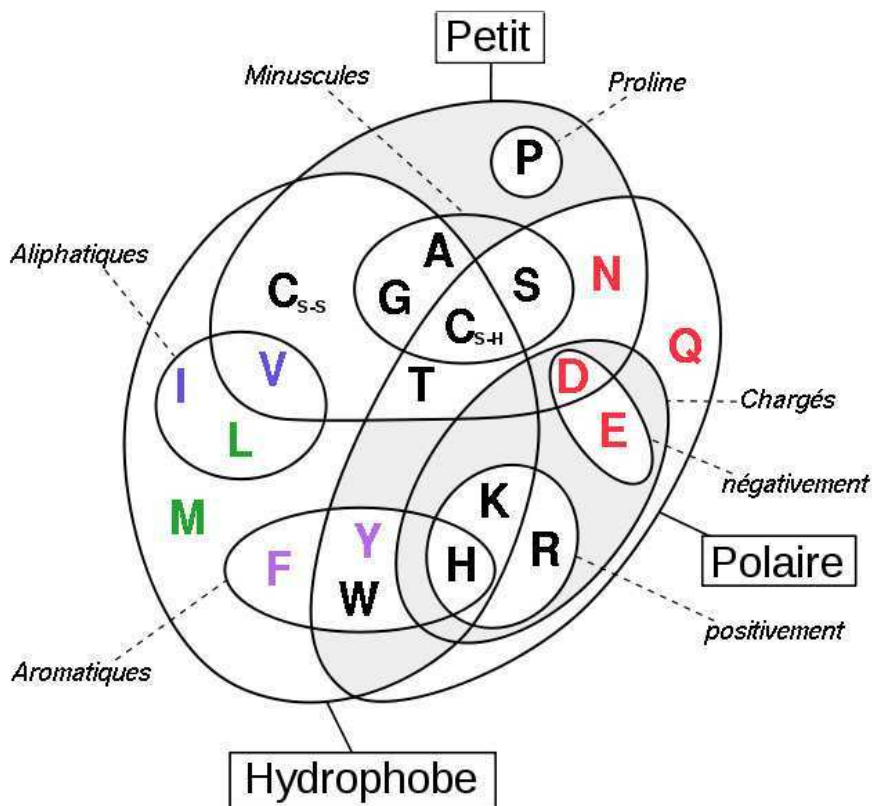
## 6. Identification de mutants par TILLING

### 6.1. Recherche

La collection de mutants EMS et azoture de Sodium de Versailles a été mise en place en collaboration avec l'URGV d'Evry dans le but de l'exploiter par génétique reverse. Cette collection a été effectuée sur environ 25000 graines de l'accession Bd21-3 de *Brachypodium distachyon*, et une population de 5628 lignées M1 a été générée. Une estimation du taux de mutation dans les lignées est de 1/520 kb. Ce chiffre, rapporté à la taille du génome (272 Mb) correspond à environ 523 mutations par plante. Si les 25 532 gènes (plus leurs promoteurs, et leurs séquences régulatrices) étaient répartis tout le long du génome, ceci correspondrait à 48 gènes mutés. Cette approximation étant très grossière, on peut considérer la présence d'une vingtaine de gènes mutés dans chaque plante. La mutagenèse par EMS et Azoture de Sodium modifie très majoritairement les bases guanines (G) en adénines (A) et les cytosines (C) en thymines (T). Considérant la redondance du code génomique sur la troisième base des codons, on considère qu'environ un tiers de ces mutations sont silencieuses. Ceci correspond alors à une quinzaine de gènes touchés par une mutation dans chaque lignée. Il est également peu probable que chacune de ces mutations ait un effet délétère sur l'activité des protéines codées par les gènes en question. En conséquent, lorsque un phénotype sera observé sur une lignée mutante pour un gène d'intérêt, il y aura une faible probabilité qu'il soit causé par une autre mutation. Cependant, la valeur du taux de mutations au départ de cette estimation a été obtenu à



**Figure 8.** Schéma de la méthode de TILLING permettant l'identification de lignées portant une mutation ponctuelle dans un gène cible.



**Figure 9.** Diagramme de Venn des acides aminés, classés par propriétés. Les acides aminés d'une même couleur peuvent plus facilement être substitués dans une séquence sans répercussions sur la structure de la protéine.

partir de statistiques faites sur certains phénotypes spécifiques et peuvent sous-estimer le nombre de mutations présentes. C'est pourquoi l'identification d'un minimum de deux allèles mutants au sein de lignées indépendantes présentant toutes les deux le même phénotype en regard du trait considéré, permettra d'écarter ce défaut dû à la mutagenèse aléatoire. Par le TILLING sur la génération M2, nous envisageons de trouver au moins cinq allèles mutants différents par gène étudié. La nature des mutations et leur localisation permettent d'obtenir des substitutions d'acides aminés dans la séquence de la protéine pouvant avoir divers degrés d'effets sur son activité. L'un des avantages d'utiliser une telle population est d'obtenir une collection d'allèles mutants plus ou moins forts permettant l'étude des conséquences de différents niveaux de variations de l'expression ou de l'activité d'un gène cible.

Grâce à la mise au point d'un système de reconnaissance par une endonucléase des mésappariements causés par une mutation ponctuelle dans les amplicons de pools d'ADN provenant de quatre plantes appartenant à huit familles indépendantes, la technique de TILLING (Targeting Induced Local Lesions in Genomes) (Fig. 9) permet d'identifier des familles de mutants portant des mutations dans des gènes cibles : les pools d'ADN génomique sont amplifiés par plusieurs cycles de PCR utilisant un couple d'oligonucléotides spécifiques du gène d'intérêt. L'oligonucléotide 5' est marqué par un chromophore d'une couleur spécifique, et l'oligonucléotide 3' est marqué par un chromophore d'une couleur différente. Après amplification, les produits de la PCR subissent plusieurs cycles de dénaturation/renaturation, et si un allèle portant une mutation ponctuelle se trouvait parmi les matrices de départ, la probabilité que les amplicons mutants soient ré-hybridés avec des amplicons sauvages est très élevée. Ceci créera des mésappariements qui seront alors reconnus et coupés par une endonucléase (endo1). Les produits PCR digérés sont alors déposés sur gel et mis à migrer, chaque pool dans un puits distinct, puis les chromophores sont révélés. Si un pool ne contenait pas de lignée mutante pour notre gène d'intérêt, aucun produit n'y aura été digéré

**Tableau 4.** Récapitulatif et description des mutations dans le gène *BdCOMT3* dans les familles identifiées par TILLING, propriétés physiques des acides aminés impliqués et scores de prédiction de tolérance de chaque substitution (scores SIFT).

Position sur l'ADNc	Position sur la protéine	Type de mutation	Famille	Propriétés physiques AA sauvage	Propriétés physiques AA muté	Scores SIFT
C629A	A210D	<b>faux-sens</b>	<b>3904</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	<b>0.01</b>
G638A	G213D	<b>faux-sens</b>	<b>211</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	<b>0.02</b>
G638A	G213D	<b>faux-sens</b>	<b>3380</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	<b>0.02</b>
G672A	G224G	silencieuse	3725	Minuscule, apolaire, hydrophobe, aliphatique		
G969A	G323G	silencieuse	3730	Minuscule, apolaire, hydrophobe, aliphatique		
G976A	E326K	<b>faux-sens</b>	<b>4142</b>	Acide (chargé -), polaire	Hydrophobe, polaire, basique (chargé +)	0.12
G721A	G241R	<b>faux-sens</b>	<b>4688</b>	Minuscule, apolaire, hydrophobe, aliphatique	Polaire, basique (chargé +)	0.24
G737A	G246D	<b>faux-sens</b>	<b>4604</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	1.00
G737A	G246D	<b>faux-sens</b>	<b>7480</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	1.00
G1063A	A355T	<b>faux-sens</b>	<b>4927</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, hydrophobe, polaire	<b>0.03</b>
C762T	P254P	silencieuse	5115	Petit, apolaire		
G767A	G256D	<b>faux-sens</b>	<b>5139</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	<b>0.00</b>
G767A	G256D	<b>faux-sens</b>	<b>7549</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	<b>0.00</b>
C498T	D166D	silencieuse	5588	Petit, acide (chargé -)		
C600T	T200T	silencieuse	5827	Petit, polaire, hydrophobe		
G616A	G206S	<b>faux-sens</b>	<b>6840</b>	Minuscule, apolaire, hydrophobe, aliphatique	Minuscule, polaire	0.76 (le seul G non conservé)
C854T	P285L	<b>faux-sens</b>	<b>7391</b>	Petit, apolaire	Apolaire, hydrophobe, aliphatique	<b>0.03</b>
G588A	E196E	silencieuse	8240	Acide (chargé -), polaire		
G708A	G236G	silencieuse	5338	Minuscule, apolaire, hydrophobe, aliphatique		
C840T	C280C	silencieuse	5348	Polaire, hydrophobe		
G37A	D13N	<b>faux-sens</b>	<b>5645</b>	Petit, acide (chargé -)	Petit, polaire	<b>0.00</b>
G117A	L39L	silencieuse	5714	Apolaire, hydrophobe, aliphatique		
G969A	G323G	silencieuse	3730	Minuscule, apolaire, hydrophobe, aliphatique		
>650pb/?		silencieuse	<b>185</b>			
G1013A	G338D	<b>faux-sens</b>	<b>5200</b>	Minuscule, apolaire, hydrophobe, aliphatique	Petit, acide (chargé -), polaire	<b>0.00</b>

et seul un amplicon de grande taille et montrant le mélange de couleur des deux chromophores sera observable sur le gel. Si au contraire le pool contenait une matrice mutante, ceci sera montré par la présence sur le gel de deux produits de taille plus réduite montrant chacun la couleur du chromophore correspondant à la partie du gène 5' ou 3'. La déconvolution, c'est-à-dire la répétition de ce processus isolément sur chaque famille du pool précédent, permet de déterminer la famille précise portant une mutation dans le gène d'intérêt, et d'identifier cette mutation par séquençage.

Notre gène candidat *BdCOMT3* a été une cible du TILLING sur les 5628 lignées (ou familles) de la collection : 25 familles de mutants ont ainsi été identifiées comme portant une mutation dans ce gène. Parmi ces mutations, dix sont silencieuses, deux n'ont pas pu être déterminées et treize sont faux-sens, parmi lesquelles trois sont présentes deux fois. Il y a donc dix allèles mutants distincts. Ce TILLING a été effectué sur le second exon du gène *BdCOMT3* à Evry par Marion Dalmais et Nicolas Oria. Les résultats sont présentés sur le Tableau 4. La comparaison des propriétés physiques des acides aminés (Fig. 9) permet d'avoir une première idée sur l'impact des mutations : des modifications de taille, de charge, ou d'hydrophobicité auront un impact plus probable que si le nouvel acide aminé a les mêmes propriétés que celui de l'allèle sauvage.

## 6.2. Prédiction des effets des mutations

### 6.2.1. Prédiction par l'outil SIFT

Sorting Intolerant From Tolerant (<http://blocks.fhcrc.org/sift/SIFT.html>) (Ng and Henikoff 2006) est un programme de prédiction de l'impact d'une substitution d'un acide aminé dans une séquence protéique, basé sur le degré de conservation des résidus dans les alignements de séquences très similaires que le programme aura identifié par PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). Si la probabilité normalisée pour une substitution est inférieure à 0,05, c'est-à-dire si la présence à l'état naturel de l'acide aminé





**Figure 10.** Séquence primaire de BdCOMT3 et de ses allèles mutants identifiés par TILLING, sur laquelle les acides aminés essentiels à l'activité de la protéine ont été indiqués: en orange les acides aminés essentiels à la dimérisation, en rouge ceux de fixation du substrat sur le site actif, en bleu clair les motifs de fixation au cofacteur, en rose les résidus catalytiques. Les acides résidus surmontés d'un tiret bleu sont particulièrement conservés dans les séquences de COMT impliquées dans la lignification. Les acides aminés modifiés sont indiqués en jaune dans les lignées mutantes et en vert sur la séquence sauvage.

**Figure 11 et 12.** Prédiction de la structure tridimensionnelle de BdCOMT3 et de ses allèles mutants, visualisées dans PyMol. Les acides aminés d'importance particulière ont été représentés d'après le même code couleur que Fig. 10: en orange les acides aminés essentiels à la dimérisation, en rouge ceux de fixation du substrat sur le site actif, en bleu clair les motifs de fixation au cofacteur, en rose les résidus catalytiques. Les acides aminés modifiés sont indiqués en jaune dans les lignées mutantes et en vert sur la séquence sauvage. Les protéines ont été orientées de façon à montrer la position du résidu muté dans chaque famille, indiqué par une flèche blanche.

La représentation en cartoon (a) et (c) montre l'assemblage des structures secondaires (hélices  $\alpha$  et feuillettes  $\beta$ ) mais ne rend pas compte de l'encombrement des résidus. La représentation en sphères (b) et (d) permet de visualiser en gros plan l'organisation dans l'espace des résidus mutés et leurs voisins.

muté est extrêmement rare à cette position dans la séquence, on considérera cette mutation comme délétère donc ayant un effet probable sur l'activité de la protéine. Les résultats de SIFT pour toutes les substitutions possibles sur l'intégralité de la séquence de BdCOMT3 sont présentés en annexe de ce manuscrit.

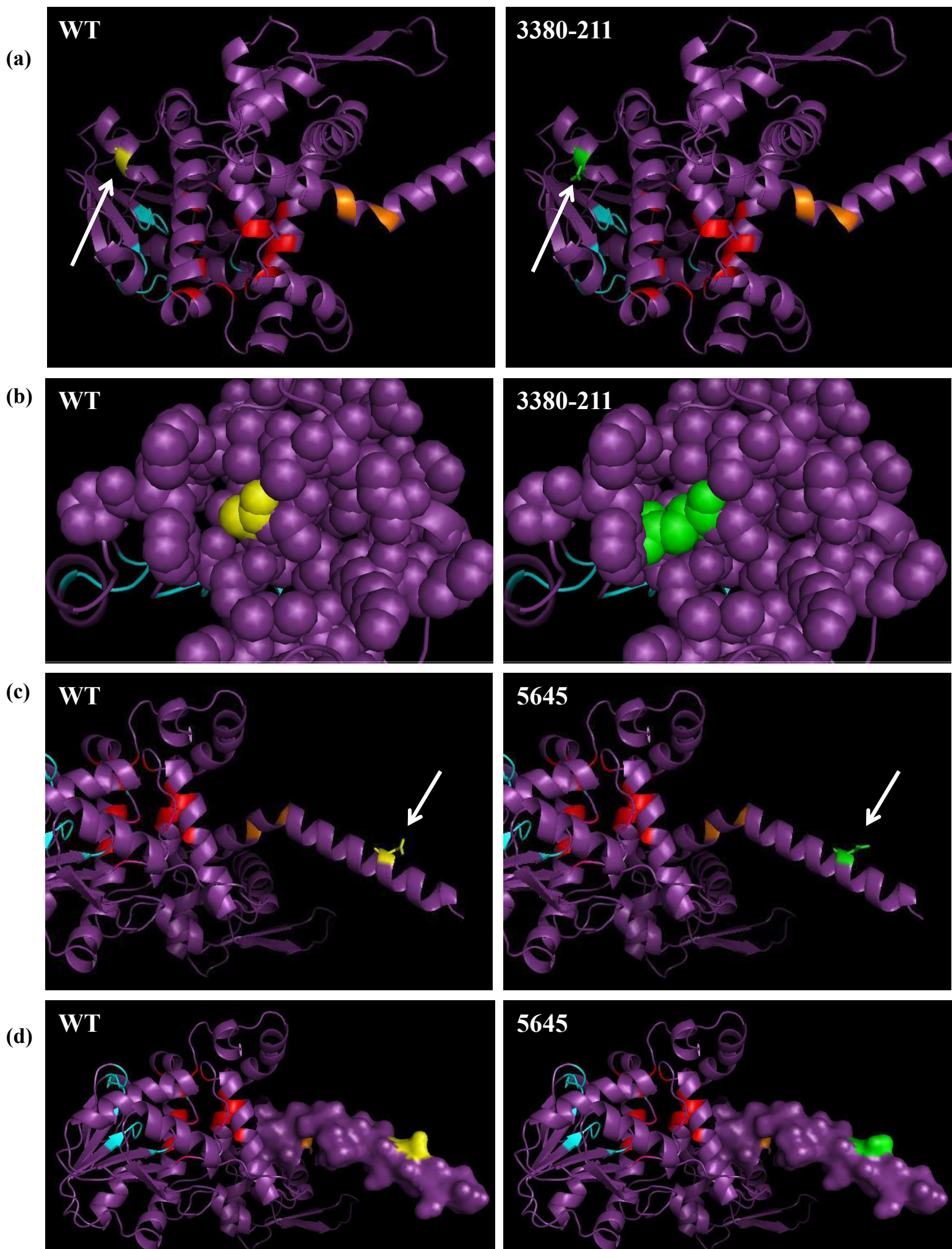
D'après l'analyse SIFT, seules 6 mutations sur les dix auront potentiellement un effet. Cependant, parmi celles-ci, deux sont situées aux extrémités de la séquence protéique : D13N à l'extrémité N-terminale (score 0.00) et A355T en C-Ter (score 0.03). Ces régions sont extrêmement variables et il est probable que sur le nombre de protéines analysées par le programme, peu ou pas d'entre elles ne présentent l'acide aminé correspondant aux allèles mutants, faisant chuter le score SIFT à des valeurs très faibles qui pourraient paraître significatives mais doivent impérativement être vérifiées. De même, des valeurs de score élevées ne sont qu'une indication qui peuvent se révéler trompeuses selon l'environnement des acides aminés considérés au sein de chaque séquence analysée par le programme.

D'après ces résultats, les lignées prioritaires pour le génotypage et l'analyse de leurs lignines sont les suivantes : 3904, 3380 (et 211), 4927, 5139 (et 7549), 7391 et 5645.

#### 6.2.2. Prédiction par modélisation

Deux facteurs ont été pris en considération, distinguant cette analyse de la précédente : la nature et les propriétés physiques des acides aminés (Tableau 4), et la position des mutations par rapport à la structure tertiaire de la protéine BdCOMT3. La Figure 10 présente la séquence primaire de BdCOMT3 sur laquelle les acides aminés essentiels à son activité ont été colorés, et les acides aminés particulièrement conservés chez les COMT d'autres espèces ont été indiqués en bleu.

Cette protéine n'ayant pas été cristallisée, il n'en existe pas de modèle tridimensionnel (3D). Il a donc été nécessaire dans un premier temps de construire une prédiction de modèle 3D de



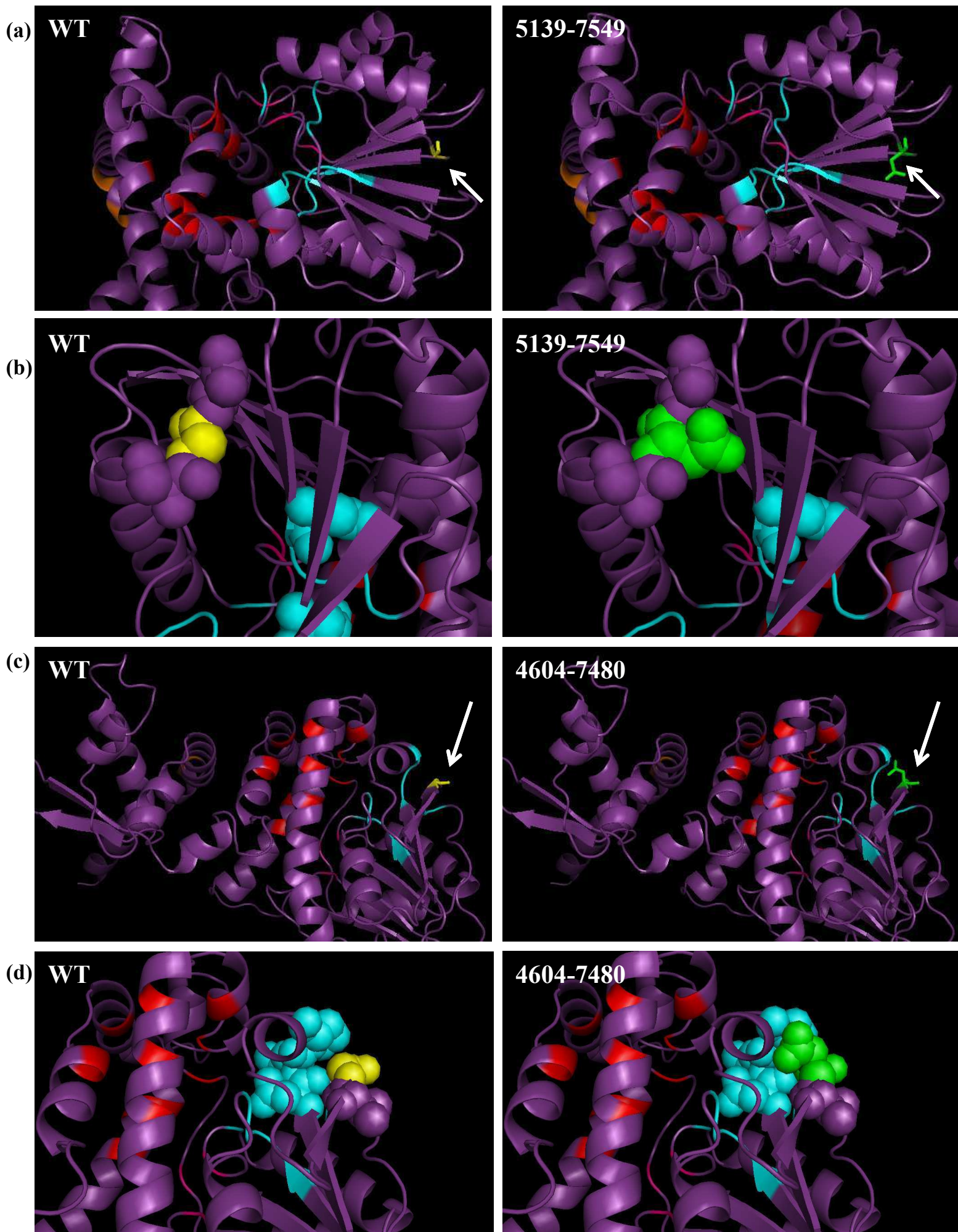
**Figure 11.** Prédiction de la structure tridimensionnelle de BdCOMT3 et de ses allèles mutants dans les familles 3380 (et 211) et 5645.

BdCOMT3. Au début de ce projet de thèse, la COMT cristallisée la plus proche de celle de *Brachypodium* était celle de luzerne (Zubieta, Kota et al. 2002) (code PDB 1kyw). Ces deux protéines présentent une identité de 62% et une similitude de 78% (acides aminés identiques ou aux propriétés physiques similaires). Récemment, la COMT de l'ivraie a été cristallisée (Louie, Bowman et al. 2010) (code PDB 3p9k). Cette protéine présente une identité de 87,4% et une similitude de 94% avec BdCOMT3 et présente l'avantage de provenir d'une graminée. Cependant, il a été montré qu'une identité de 30% permet de construire des modèles tridimensionnels robustes permettant l'analyse des effets de mutagenèses dirigées (Baker and Sali 2001). Les prédictions générées à partir d'une matrice ou de l'autre sont toutes les deux identiques et satisfaisantes pour l'utilisation qui en sera faite. Les prédictions par homologie de modèles de la protéine sauvage ont été générées sur le serveur Geno3D, puis visualisés avec le programme Pymol, celui-ci permettant de modifier ponctuellement un acide aminé. Les modélisations ont été représentées en vue d'ensemble sous forme de « cartoon », c'est-à-dire montrant les structures secondaires (hélices  $\alpha$  et feuilletts  $\beta$ ) pour situer la mutation sur la structure 3D, puis en gros plan avec une représentation en « sphères » montrant l'encombrement de chaque acide aminé ou en « surface » si la modification n'avait d'impact que sur la périphérie de la protéine (Fig. 11 et 12).

Deux lignées attirent particulièrement l'attention sur la séquence primaire : la 6840 présente une substitution au sein du motif de Rossmann, indispensable à la fixation du cofacteur SAM, et la lignée 4142 présente une substitution sur un résidu catalytique ce qui suggère une forte diminution de l'activité de la protéine.

Les lignées décrites ici sont celles dont des individus homozygotes pour la mutation ont par la suite été identifiées puis analysés.

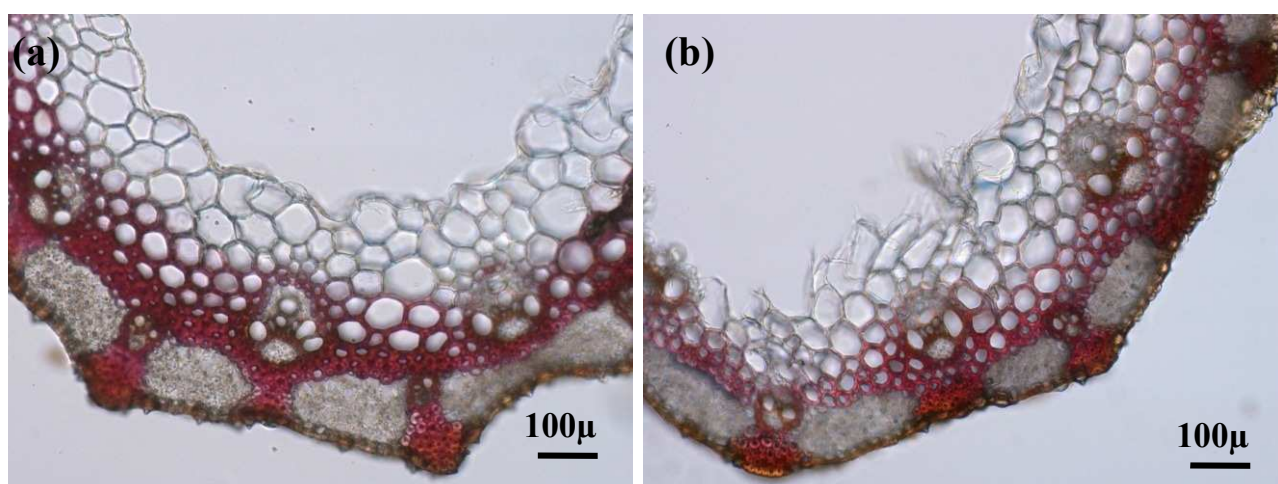
- 3380 (Fig. 11): La substitution d'une glycine en acide aspartique est très fréquente lors de la mutagenèse chimique par EMS ou Azoture de sodium. Elle a pour conséquence



**Figure 12.** Prédiction de la structure tridimensionnelle de BdCOMT3 et de ses allèles mutants dans les familles 5139 (et 7549) et 4604 (et 7480).

une modification de taille, de charge et d'hydrophobicité, ce qui peut potentiellement avoir de lourdes conséquences sur une protéine, surtout lorsque la mutation est située à un résidu clé de l'activité. Dans le cas de la lignée 3380, la substitution est située à 6 acides aminés du motif de Rossmann DVDGGGV, ce qui est relativement éloigné. Considérant les repliements de la séquence tertiaire, l'encombrement créé par l'acide aspartique pourrait modifier la position du feuillet  $\beta$  situé en face de sa chaîne latérale, dont l'extrémité porte un résidu impliqué dans la fixation du cofacteur. Si la modification empêche la fixation du SAM, cela pourrait avoir un effet sur l'activité de la protéine.

- 5645 (Fig. 11) : Bien que la substitution dans cette famille ait lieu au niveau du site de dimérisation, le changement d'un acide aspartique par une asparagine ne modifie que la charge et la polarité et ne crée pas de différence d'encombrement stérique. Il semble peu probable qu'elle ait un effet important, et ce malgré son score SIFT.
- 5139 (Fig. 12): La mutation a lieu en périphérie extérieure de la protéine à l'opposé du site de dimérisation. Le seul effet envisageable de cette mutation serait qu'à cause de la modification de charge et l'encombrement causé par la taille de l'acide aspartique, l'hélice  $\alpha$  située en face du résidu muté soit déséquilibrée. L'extrémité de cette structure secondaire portant un résidu impliqué dans la fixation du cofacteur SAM, l'ancrage de ce dernier et le transfert de méthyle pourrait être modifié ou empêché, ceci ayant de graves conséquences sur l'activité de BdCOMT3. Cette lignée avait un excellent score SIFT (0.00) et la modélisation ne contredit pas ce résultat.
- 4604 : La mutation a également lieu en périphérie extérieure de la protéine à l'opposé du site de dimérisation, mais précisément en face d'un groupe de résidus impliqués dans la fixation du cofacteur. L'acide aspartique étant à la fois plus gros que la glycine, il se retrouve en contact avec le site de fixation du cofacteur. Ce contact, ajouté à la charge



**Figure 13.** Coupes transversales d'entre-nœuds secs de Wt (a) et du mutant *Bdcomt3* homozygote issu de la lignée 5645 (b) colorés au Mäule. Les unités S de la lignine se colorent en rouge : elles sont majoritaires au niveau des fibres interfasciculaires et absentes au niveau des faisceaux vasculaires. Les unités G sont colorées en marron et sont majoritaires dans les faisceaux vasculaires. Aucune différence de coloration n'est observée entre le sauvage et la lignée 5645.

**Tableau 5.** Etude des lignines des tiges matures de la lignée homozygote mutante 5645 par thioacidolyse sur matériel non extrait. Les résultats correspondent aux valeurs moyennes et aux erreurs moyennes entre doubles. Une seule plante mutante a été analysée. Le rendement total en monomères (H+G+S) de thioacidolyse est exprimé en μmoles par gramme de matériel végétal analysé. Les acides p-coumariques PC et féruliques FE libérés par thioacidolyse sont exprimés en mg/g de matériel végétal.

	S/G	%H	%G	%S	Rendement (μmol/g matériel végétal)			
					total	5-OHG	PC (mg/g)	FE (mg/g)
WT	2.77 (± 0.08)	2.5 (± 0.05)	26.0 (± 0.6)	71.5 (± 0.65)	179	1	7.65 (± 0.38)	5.36 (± 0.335)
5645	2.63	3.2	26.7	70.1	172	0.5	9.36	6.60

négative du nouvel acide aminé pourrait modifier l'organisation de ce site et gêner la fixation du cofacteur SAM, diminuant ainsi l'activité de BdCOMT3. Cette lignée reste prioritaire, en accord avec les résultats SIFT.

## 7. Etude de la lignification des mutants

Les mutations ponctuelles touchant le gène cible *BdCOMT3* étant vraisemblablement récessives, des plantes de lignée M2 des familles identifiées par TILLING ont été génotypées afin de caractériser des individus portant chaque mutation à l'état homozygote. Les lignines de lignées mutantes pour le gène *COMT* possèdent des signatures spécifiques en thioacidolyse. Ces signatures ont alors été recherchées dans les lignées TILLING. Celles-ci ne présentaient pas de phénotype visuel, différence de taille ou de couleur par rapport au sauvage.

La caractéristique majeure des lignines de mutants *comt* étant un effondrement de la fréquence des unités S, des coupes transversales de la première lignée mutante homozygote identifiée, la 5645, ont été observées après coloration Mäule (Fig. 13). Cette méthode colorant les unités S en rouge et les unités G en marron permet de détecter les variations importantes de présence des unités S dans les tissus lignifiés. Comme cette coloration ne révèle pas de différence visible entre témoin et mutant du point de vue de la répartition des unités S et G (Fig. 13), cela signifie que la mutation de la lignée 5645 n'affecte pas la lignification ou bien qu'elle l'affecte trop peu pour que ce soit visible par histochimie. La légère différence d'intensité de la coloration peut provenir de l'épaisseur de la paroi des fibres, qui sont plus fines chez le mutant, vraisemblablement dû à un retard de croissance causé par la mutagenèse. Cependant, cette famille portant une mutation pour laquelle un effet important n'était pas prédit, il est possible que celui-ci soit trop peu marqué pour être observé par coloration, et ses lignines ont alors été analysées par thioacidolyse (Tableau 5). Celles-ci n'ont pas présenté de différences



**Tableau 6.** Etude des lignines des tiges matures des lignées homozygotes mutantes 3380, 5139, 4604 et 5200. Les résultats correspondent aux valeurs moyennes et aux erreurs moyennes entre doubles. Les valeurs significativement différentes entre le sauvage et les mutants sont indiquées en gras. Le taux de lignine Klason (%LK) est exprimé en % pondéral du résidu pariétal (échantillon débarrassé de ses composés solubles). Le rendement total en monomères de thioacidolyse (rdt) est exprimé en  $\mu$ moles par gramme de LK. Les acides p-coumariques PC et féruliques FE libérés par thioacidolyse sont exprimés en mg/g de RP. Les deux parties du tableaux correspondent à deux lots de cultures différents, accompagnés de leurs témoins Bd21-3 respectifs.

	%RP	%LK	S/G	%H	%G	%S	Rendement ( $\mu$ mol/gLK)		PC (mg/g)	FE (mg/g)
							total	5-OHG		
WT Bd21-3	79.6 ( $\pm$ 0.6)	17.82 ( $\pm$ 0.385)	2.14 ( $\pm$ 0.09)	3.5 ( $\pm$ 0.15)	31 ( $\pm$ 1)	66 ( $\pm$ 1)	1215 ( $\pm$ 16)	3.1 ( $\pm$ 0.25)	6.20 ( $\pm$ 0.015)	9.28 ( $\pm$ 0.375)
3380	77.9 ( $\pm$ 0.8)	18.73 ( $\pm$ 0.06)	2.23 ( $\pm$ 0.06)	3.4 ( $\pm$ 0.05)	30 ( $\pm$ 0.5)	67 ( $\pm$ 0.5)	1303 ( $\pm$ 30.5)	3.8 ( $\pm$ 0.25)	6.46 ( $\pm$ 0.08)	9.16 ( $\pm$ 0.105)
5139	80.5	<b>14.81</b> ( $\pm$ <b>0.01</b> )	<b>1.13</b> ( $\pm$ <b>0.00</b> )	3.7 ( $\pm$ 0.05)	45 ( $\pm$ 0.00)	51 ( $\pm$ 0.00)	<b>925</b> ( $\pm$ <b>4.5</b> )	<b>57.0</b> ( $\pm$ <b>0.105</b> )	<b>5.09</b> ( $\pm$ <b>0.02</b> )	7.91 ( $\pm$ 0.08)
4604	80.4	16.96 ( $\pm$ 0.005)	<b>1.44</b> ( $\pm$ <b>0.01</b> )	3.3 ( $\pm$ 0.00)	40 ( $\pm$ 0.00)	57 ( $\pm$ 0.00)	<b>1132</b> ( $\pm$ <b>16</b> )	<b>38.4</b> ( $\pm$ <b>0.6</b> )	6.32 ( $\pm$ 0.05)	8.47 ( $\pm$ 0.02)
WT Bd21-3	77.8	17.45 ( $\pm$ 0.03)	2.03 ( $\pm$ 0.01)	3.4 ( $\pm$ 0.00)	32 ( $\pm$ 0.00)	65 ( $\pm$ 0.00)	1294 ( $\pm$ 6)	2.91 ( $\pm$ 0.045)	6.25 ( $\pm$ 0.00)	9.34 ( $\pm$ 0.06)
5200	80.6	15.89 ( $\pm$ 0.05)	<b>1.46</b> ( $\pm$ <b>0.01</b> )	<b>5.1</b> ( $\pm$ <b>0.05</b> )	39 ( $\pm$ 0.00)	56 ( $\pm$ 0.00)	<b>927</b> ( $\pm$ <b>12.5</b> )	2.47 ( $\pm$ 0.01)	7.54 ( $\pm$ 0.12)	6.74 ( $\pm$ 0.285)

significatives avec le sauvage, tant au niveau des quantités d'acides hydroxycinnamiques et de monomères inhabituels (5-OH-G) qu'au niveau des proportions d'unités incorporées : le rapport S/G, les pourcentages de S, G et H ainsi que le rendement de thioacidolyse ne varient pas significativement. La mutation portée par la lignée 5645 (G37A traduite D13N) ne semble pas avoir d'effet sur l'activité de la protéine BdCOMT3.

Les lignées identifiées par la suite ont directement été soumises à l'étude de leurs lignines en termes de quantité (dosage de la teneur en lignine Klason) et de structure (thioacidolyse) (Tableau 6):

Les lignines du mutant 3380 ne présentent pas de différences significatives avec le sauvage, contrairement à ce qui avait été prédit précédemment.

Les lignées 4604, et de façon plus marquée encore 5139, présentent une diminution du taux de lignine de leurs tiges matures (elles perdent respectivement un et trois points de lignine Klason, lorsque ce taux est exprimé en % pondéral du résidu pariétal RP, c'est-à-dire de la paille débarrassée des composés solubles), un rendement de thioacidolyse plus faible indiquant une augmentation de la proportion en liaisons résistantes entre les unités du polymère, un rapport S/G plus faible que le sauvage, et la présence du marqueur 5-OH-G en quantité plus de dix fois plus importante que chez le sauvage. De plus, la quantité d'acide *p*-coumarique libéré par thioacidolyse est diminuée par la mutation, ce qui suggère qu'il y a moins d'esters *p*-coumariques associés aux parois de ce mutant. Dans le cas du maïs, les travaux de Ralph et al ont montré que cet ester est surtout associé aux unités S. Chez la lignée 5139, cette perte d'acide *p*-coumarique associé aux parois peut donc être majoritairement imputable à la plus faible fréquence en unités S. Ces lignées mutantes présentent donc toutes les caractéristiques de graminées déficientes pour la COMT tels que les mutants naturels de maïs *bm3* (Piquemal, Chamayou et al. 2002) ou les mutants de sorgho *bmr12* (Bout and Vermerris 2003) mais avec un phénotype légèrement moins marqué au niveau des lignines et des composés phénoliques.



Une autre différence est l'absence de phénotype de couleur (quand *bm3* et *bmr12* ont précisément été identifiées grâce à la coloration de leurs tissus lignifiés). Ces différences ne constituent que des variations d'intensité des phénotypes, et ne modifient pas leur nature. Elles proviennent vraisemblablement du fait que les mutations ponctuelles portées par les lignées 4604 et 5139 n'inactivent pas entièrement *BdCOMT3* mais ne font qu'en diminuer l'efficacité. La lignée 5200 possède des lignines différentes de celles du sauvage avec un rendement de thioacidolyse et un rapport S/G plus faibles et similaires à ceux des mutants décrits précédemment. Cependant la quantité du marqueur de déficience *comt*, le monomère 5-OH-G, n'est pas modifiée. Ces caractéristiques particulières suggèrent que les modifications de lignines de cette lignée proviennent de la perte de fonction d'un autre gène conjointement ou non à une perte de fonction *BdCOMT3*.

## 8. Conclusion et perspectives

La convergence des données phylogénétiques, d'expression et fonctionnelles suggère le rôle de *BdCOMT3* dans la voie de biosynthèse des monolignols chez *Brachypodium distachyon*. Le reste des lignées identifiées par TILLING est en cours de génotypage afin de caractériser un plus grand nombre de lignées possédant des allèles différents affectant l'activité de la protéine et la composition en lignine des parois. Parmi celles-ci, se trouvent des lignées pour lesquelles la mutation est prédite comme ayant un effet délétère sur l'activité de la protéine. Il pourra ensuite être vérifié si les lignines modifiées ont un impact sur les rendements de saccharification des parois. En effet, l'une des caractéristiques des mutants *comt* est la diminution de la quantité d'acide *p*-coumarique dans leur paroi (Kuc, Nelson et al. 1968). Or ce composé joue un rôle dans la liaison des lignines à la paroi et constitue l'une des cibles de l'amélioration des parois végétales à destination de la production de biocarburants (Simmons, Loqué et al. 2010).



## Chapitre IV :

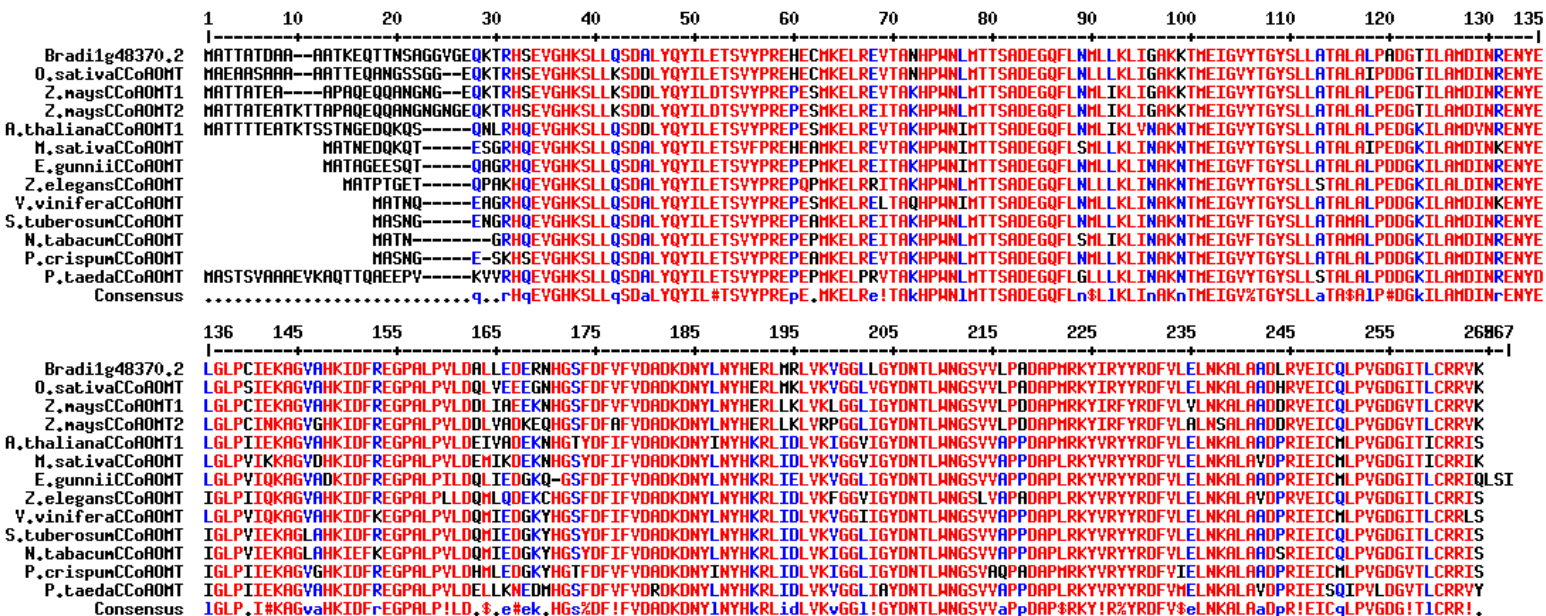
Identification de la  
cafféoyl-coenzyme A-3-*O*-méthyltransférase impliquée  
dans la voie de biosynthèse des monolignols

## Chapitre IV : Identification de la cafféoyl-coenzyme A 3-*O*-méthyltransférase impliquée dans la voie de biosynthèse des monolignols

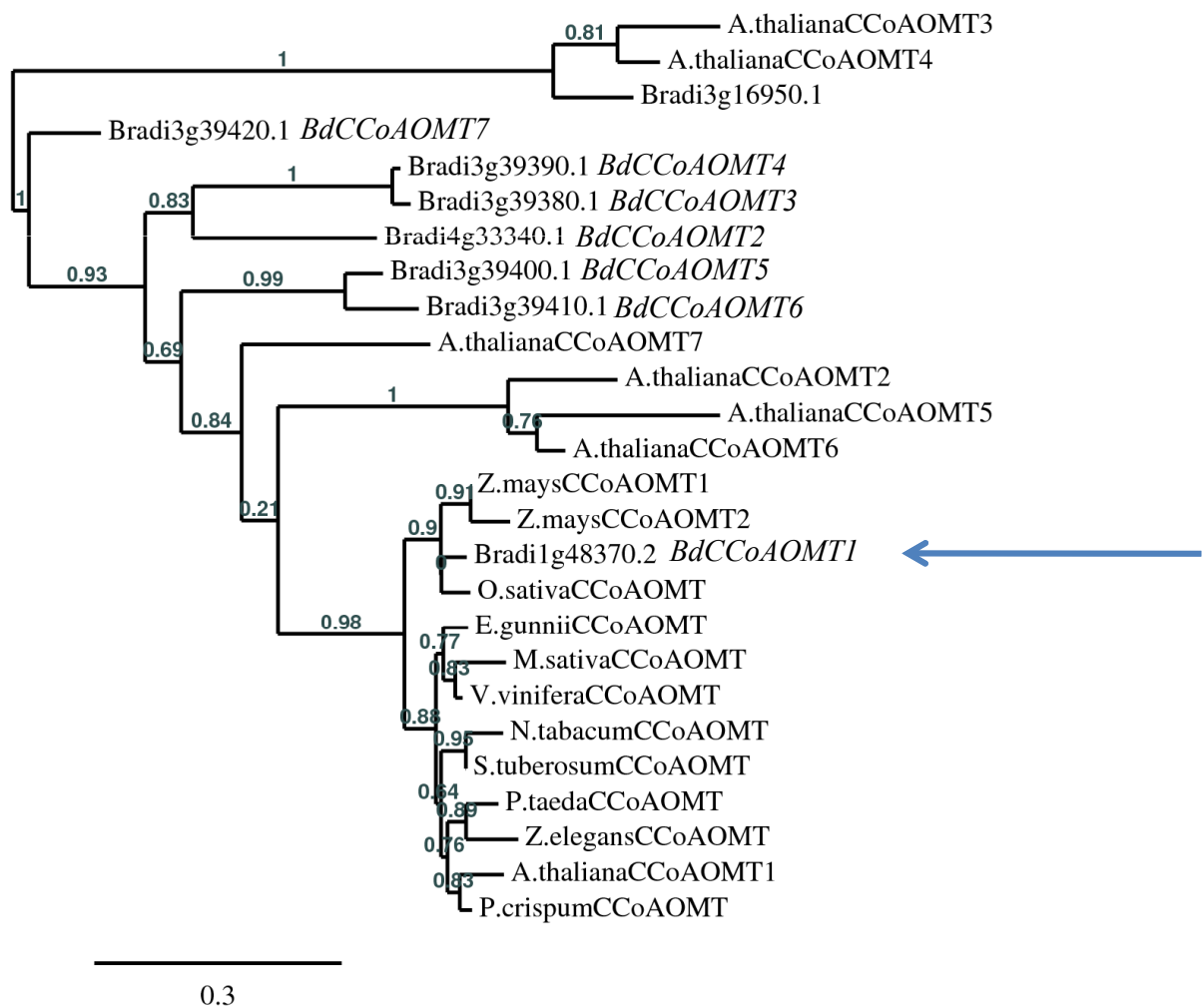
Ce chapitre présente la détermination chez *Brachypodium distachyon* du gène candidat pour la cafféoyl-coenzyme A 3-*O*-méthyltransférase (CCoAOMT) impliquée dans la voie de biosynthèse des monolignols, la recherche de mutants dans ce gène et leur caractérisation.







**Figure 2.** Alignement du candidat CCoAOMT Bradi1g48370.2 avec des CCoAOMT impliquées dans la lignification chez d'autres espèces végétales. La couleur rouge montrant les forts consensus, on distingue la grande similarité de séquence entre ces protéines.



**Figure 3.** Arbre phylogénétique construit à partir des séquences protéiques de CCoAOMT identifiées dans plusieurs espèces, incluant celles codées par les gènes candidats de *Brachypodium distachyon*. Les chiffres gris indiquent la valeur « bootstrap » représentant la probabilité d'existence de chaque branche du phylogramme lors des répétitions des calculs de construction. La longueur des branches est proportionnelle à la distance évolutive entre les nœuds. La flèche bleue indique le gène candidat de *Brachypodium* retenu dans cette étude.

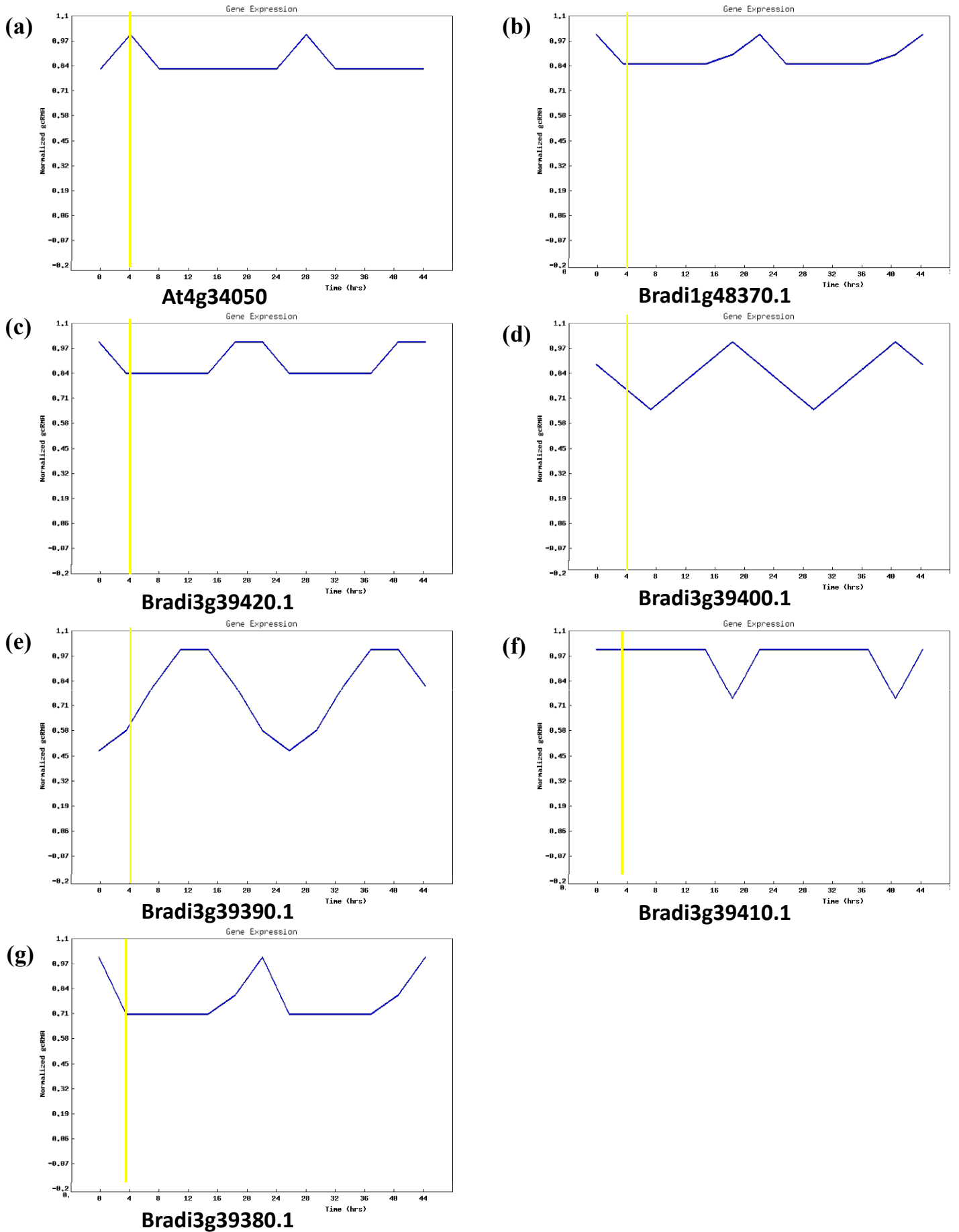
La cafféoyl-coenzyme A 3-*O*-méthyltransférase est impliquée dans la première étape de méthylation de la voie de biosynthèse des monolignols où elle méthyle le cycle aromatique du cafféoyl-CoA et du 5-hydroxyfêruloyl-CoA, produisant respectivement le fêruloyl-CoA et le sinapoyl-CoA (Parvathi, Chen et al. 2001).

#### 1. Détermination du gène candidat par alignements de séquences

Parmi les gènes *CCoAOMT* dont l'implication dans la voie de biosynthèse des monolignols a été montrée, la traduction en séquence protéique de la CCoAOMT1 d'*Arabidopsis thaliana* (*AtCCoAOMT1* Locus: At4g34050) a été choisie pour servir d'entrée au BLAST sur le génome 4X, puis 8X de *Brachypodium distachyon*. Parmi les gènes de *Brachypodium* issus de cette recherche, ceux possédant les scores les plus élevés et les e-value les plus faibles sont les suivants :

	Score	P(N)
Bradi1g48370.2	1102	1.1e-112
Bradi1g48370.1	651	5.5e-111
Bradi3g39420.1	722	2.0e-72
Bradi3g39400.1	716	8.6e-72
Bradi3g39390.1	684	2.1e-68
Bradi4g33340.1	663	3.6e-66
Bradi3g39410.1	662	4.6e-66
Bradi3g39380.1	658	1.2e-65
Bradi3g16950.1	343	2.9e-32

Le BLAST à partir de séquences de CCoAOMT d'autres espèces donne les mêmes résultats, seuls les scores varient. Il est à noter que Bradi1g48370.2 et Bradi1g48370.1 correspondent au même gène transcrit sous deux formes après épissage alternatif. Nous ne tiendrons compte que de Bradi1g48370.2 qui présente un exon en moins et dont le score d'alignement est supérieur.



**Figure 4.** Expression en fonction du cycle circadien des gènes *AtCCoAOMT1* (a), *BdCCoAOMT7* (b), *BdCCoAOMT5* (c) et *BdCCoAOMT4* (d), *BdCCoAOMT6* (e), *BdCCoAOMT4* (f), *BdCCoAOMT3* (g) et dans les conditions de culture les plus proches de celles de Versailles (Pas de données disponibles sur *BdCCoAOMT2*).

Les séquences protéiques de ces huit gènes de *Brachypodium* ont été alignées sur celles d'autres CCoAOMT (Fig.1). Dix huit acides aminés nécessaires à l'activité de la protéine (Dixon et al, 2005) ont été observés sur cet alignement où ils sont désignés par une flèche jaune (Fig. 1). Ils sont bien conservés ou, dans certains cas, sont substitués par des acides aminés aux propriétés similaires. Bradi3g39380.1 et Bradi3g39390.1 présentent des différences importantes pour trois de ces résidus: en positions 79, 201 et 202 de l'alignement. De même, les séquences de Bradi3g39420.1 et Bradi3g39410.1 ne possèdent pas les acides aminés habituellement conservés à ces deux dernières positions qui sont PA ou LA.

Les résidus des séquences de Bradi3g39400.1 et Bradi4g33340.1 ne sont pas conservés par rapport aux autres CCoAOMT aux positions 303, 305 et 309 de l'alignement. La dernière protéine issue du BLAST, Bradi3g16950.1, semble plus proche d'AtCCoAOMT3 et AtCCoAOMT4 que des autres protéines alignées et est très différente des sept autres protéines de *Brachypodium*. L'alignement de Bradi1g48370.2 sur les CCoAOMT des autres espèces végétales (Fig.2) montre une très grande similarité de séquence, soulignée par la quantité de résidus à fort consensus (en rouge).

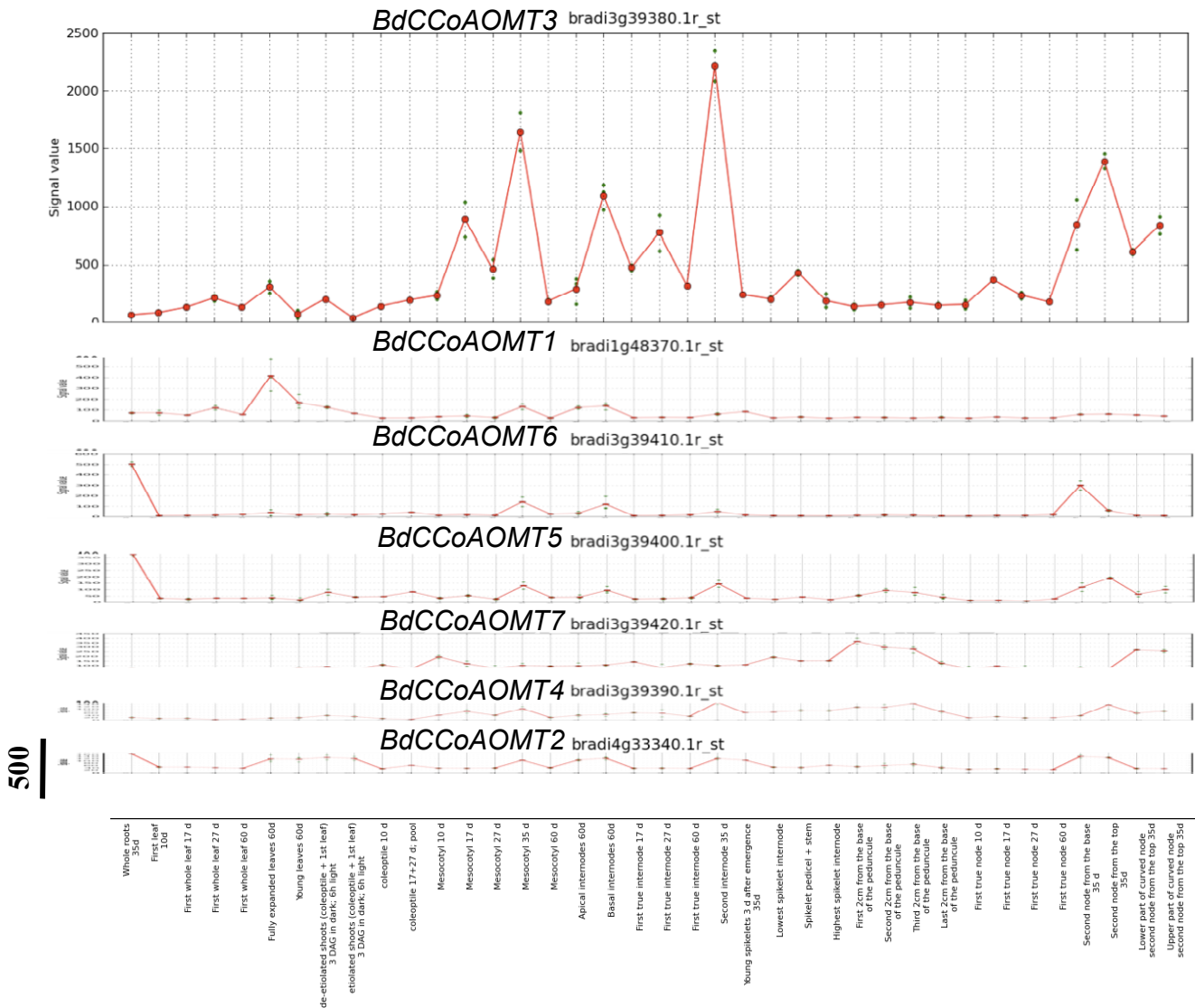
Le nombre d'EST a été cherché pour chacun de ces gènes (Tableau 1) et indique que le plus exprimé serait Bradi3g39420.1 (avec 20 EST) suivi de Bradi1g48370.2, avec 3 EST. Ceci, ajouté à son haut score de BLAST et aux observations de séquences, suggère le gène Bradi1g48370.2 comme principal candidat *CCoAOMT* dans un premier temps.

## 2. Etude phylogénétique

Un arbre phylogénétique a été créé à partir des séquences alignées précédemment (Fig. 3). Les CCoAOMT identifiées chez les graminées utilisées pour cette analyse (*Zea mays*, et *Oryza sativa*) sont regroupées au sein d'un sous-groupe contenant Bradi1g48370.2 (BdCCoAOMT1). Toutes les autres protéines de *Brachypodium* présentes sur cet arbre sont situées en dehors du

**Tableau 1.** Correspondance entre les appellations des gènes *CCoAOMT* annotés chez *Brachypodium distachyon*.

Nom d'après leur position sur le génome	Nomenclature simplifiée	EST
<a href="#">Bradi1g48370.2</a>	BdCCoAOMT1	3
<a href="#">Bradi4g33340.1</a>	BdCCoAOMT2	0
<a href="#">Bradi3g39380.1</a>	BdCCoAOMT3	2
<a href="#">Bradi3g39390.1</a>	BdCCoAOMT4	2
<a href="#">Bradi3g39400.1</a>	BdCCoAOMT5	0
<a href="#">Bradi3g39410.1</a>	BdCCoAOMT6	1
<a href="#">Bradi3g39420.1</a>	BdCCoAOMT7	20



**Figure 5.** Niveaux d'expression des trois *CCoAOMT* annotées de *Brachypodium* dans des organes différents. Données provenant de Bradinet. Afin de faciliter les comparaisons, les trois graphes sont à la même échelle en ordonnée (échelle indiquant 500 unités d'expression). Les différents tissus sont indiqués en abscisse.

cluster regroupant toutes les CCoAOMT impliquées dans la lignification, confirmant le choix de *BdCCoAOMT2* comme unique candidat de *CCoAOMT* impliquée dans la lignification chez *Brachypodium distachyon*.

### 3. Expression en fonction du cycle circadien

Seule *BdCCoAOMT6* présente une expression maximale après 4h de jour, mais pas sous forme de pic comme dans le cas d'*AtCCoAOMT1*. Comme pour les gènes précédents, les données obtenues sur Diurnal (Fig. 4) ne permettent pas de conclure pour les candidats *CCoAOMT* de *Brachypodium* à une expression d'un gène spécifique des lignines du type de celui observé chez *Arabidopsis thaliana*, présentant un pic d'expression au bout de quatre heures d'exposition à la lumière.

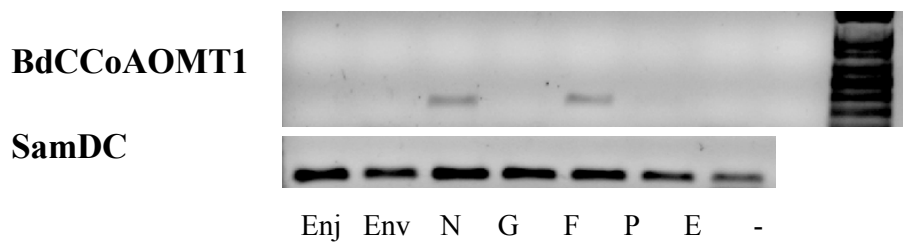
### 4. Profils d'expression en fonction des tissus

#### 4.1. RT-PCR

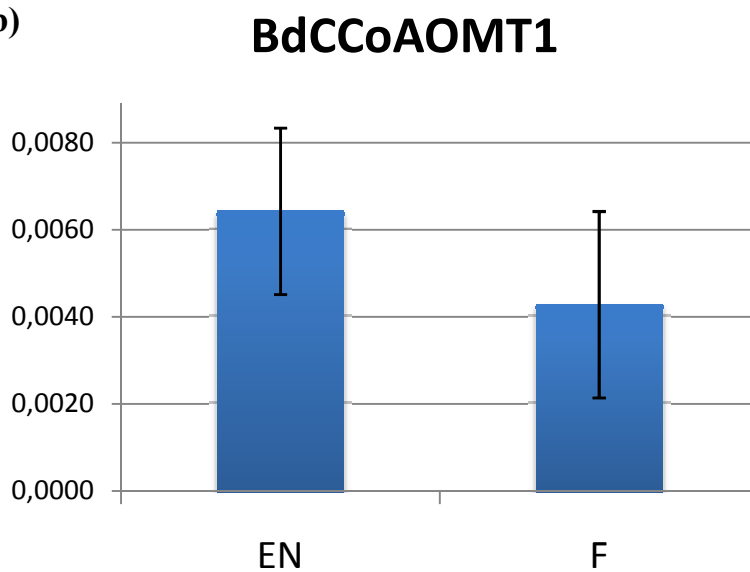
L'amplification du gène *BdCCoAOMT1* s'est révélée difficile. En RT-PCR semi-quantitative, aucune donnée d'expression du gène n'a pu être obtenue pour les tiges collectées aux quatre stades analysés au cours de ces travaux. L'étude sur les différents organes au stade reproductif (Fig. 5a) pourrait suggérer une expression légèrement plus importante dans les feuilles et dans les nœuds. L'analyse par RT-PCR quantitative sur entre-nœuds et feuilles (Fig. 5b) indique une expression très faible, sans différence significative entre les deux organes.

Malgré ces résultats peu probants, les indications issues de l'analyse phylogénétique nous ont poussé à continuer le travail sur ce gène

(a)



(b)



**Figure 5.** (a) RT-PCR au stade 7 semaines sur différents organes du maître brin de trois plantes. Enj: jeunes entre-nœuds (en cours de lignification); Env: entre-nœuds lignifiés; N: nœuds; G: gaine foliaire; F: feuilles; P: pédoncule; E: épi. L'amplification du gène constitutif SamDC permet d'indiquer la quantité de matrice présente dans chaque puits. (b) RT-PCR quantitative montrant les niveaux d'amplification de BdCOMT3 rapportés à ceux du gène de ménage SamDC dans les entre-nœuds (EN) et les feuilles (F) de quatre répétitions biologiques de Bd21- 3 après émergence de l'épi. Les données ont été normalisées par rapport au gène constitutif SamDC.

## 4.2. Profils Bradinet

D'après les données issues du site Bradinet, la *CCoAOMT* la plus exprimée dans les tissus lignifiés, en particulier les entre-nœuds après épiaison, serait *BdCCoAOMT3*. Ceci suggère que, contrairement à ce qu'indiquaient les observations de séquences et d'EST, *BdCCoAOMT3* serait le candidat principal pour la voie de biosynthèse des monolignols. *BdCCoAOMT1* n'est que faiblement exprimé et son expression est principalement localisée dans les feuilles. Ceci explique les difficultés rencontrées lors de l'amplification de ce gène par RT-PCR.

Cependant, les données de Bradinet n'ayant été disponibles que récemment, les travaux présentés ici avaient été menés sur le candidat *BdCCoAOMT1* qui semblait le plus prometteur.

## 5. Complémentation fonctionnelle

La perte de fonction de *CCoAOMT1* chez *Arabidopsis thaliana* conduit principalement à une diminution du taux de lignine, déterminée par la méthode de Klason, et à une augmentation du rapport S/G liée à un déficit d'unités G (Do et al., 2007). La quantité de matériel nécessaire au dosage Klason n'étant pas disponible sur les plantes de lignées T1 qui ne peuvent pas être poolées, seule la thioacidolyse a été réalisée. Les résultats obtenus confirment que le mutant *ccoamt-1* a un rapport S/G plus élevé que l'échantillon témoin WS (Tableau 2).

La thioacidolyse sur la première génération de transformants (T1) montre que la baisse de rendement (donc de quantité d'unités liées uniquement en  $\beta$ -O-4, quantité pas nécessairement représentative de la quantité totale de lignine contenue dans les plantes) observée chez le mutant par rapport au sauvage n'est pas modifiée dans les quatre lignées *Atccomt-1* dans lesquelles *BdCCoAOMT1* est exprimé sous contrôle du promoteur 35S. Au lieu d'être augmentée, elle est même légèrement diminuée dans trois de ces lignées. Par ailleurs, on constate que la fréquence molaire des monomères H (%H) qui sont obtenus en faible quantité à



**Tableau 2.** Etude par thioacidolyse sur matériel non extrait des lignines des tiges matures de la première génération (T1) de lignées *d'Arabidopsis thaliana* mutantes pour le gène *CCoAOMT1* exprimant le gène *BdCCoAOMT1* (*ccoamtCp*). Une seule plante de chaque lignée complémentée a été analysée, et quatre plantes pour le témoin sauvage WS et pour la lignée mutante EBL87 ont été réunies et analysées en tant qu'échantillon. Le rendement total en monomères (H+G+S) de thioacidolyse est exprimé en  $\mu$ moles par gramme de matériel végétal sec analysé.

	Rendement en $\mu$ moles/g				Valeurs relatives			
	H	G	S	Total	S/G	%H	%G	%S
Col0	0,8 $\pm$ 0,1	95,5 $\pm$ 1	44 $\pm$ 1	139 $\pm$ 1	0,45 $\pm$ 0,00	0,58 $\pm$ 0,04	68,5 $\pm$ 0,1	30,9 $\pm$ 0,0
<i>ccoamt-1</i>	1,66	62,02	44,57	108,25	0,72	1,54	57,29	41,17
col0+Bd1 <i>ccoamtCp5</i>	1,84	80,88	44,78	127,49	0,55	1,44	63,44	35,12
<i>ccoamt-1</i> +Bd1 <i>ccoamtCp 8</i>	1,12	57,31	37,33	95,76	0,65	1,17	59,84	38,98
<i>ccoamt-1</i> +Bd1 <i>ccoamtCp 10</i>	1,22	56,97	33,88	92,07	0,59	1,33	61,87	36,80
<i>ccoamt-1</i> +Bd1 <i>ccoamtCp 14</i>	1,73	66,88	42,07	110,67	0,63	1,56	60,43	38,01
<i>ccoamt-1</i> +Bd1 <i>ccoamtCp 22</i>	1,57	62,44	39,40	103,40	0,63	1,52	60,38	38,10

**Tableau 3.** Etude des lignines des tiges matures de la troisième génération (T3) de lignées *ccoamt-1* complémentées (*ccoamtCp*). Les résultats correspondent aux valeurs moyennes et aux erreurs moyennes entre doubles. Le taux de lignine Klason (%LK) est exprimé en % pondéral du résidu pariétal (échantillon débarrassé de ses composés solubles). Le rendement total en monomères de thioacidolyse est exprimé en  $\mu$ moles par gramme de LK.

	% LK	Rendement de thioacidolyse en $\mu$ mole /g LK				S/G
		H	G	S	total	
Col0	20,23 $\pm$ 0,07	4,3 $\pm$ 0,00	848 $\pm$ 16	286 $\pm$ 8	1138 $\pm$ 24	0,34 $\pm$ 0,01
<i>ccoamt-1</i>	15,33 $\pm$ 0,01	6,8 $\pm$ 0,5	765 $\pm$ 15	367 $\pm$ 15	1139 $\pm$ 31	0,48 $\pm$ 0,01
<i>ccoamtCp 22-2</i>	14,56 $\pm$ 0,04	6,1 $\pm$ 0,3	667 $\pm$ 33	294 $\pm$ 22	966 $\pm$ 55	0,44 $\pm$ 0,01
<i>ccoamtCp 8-1</i>	15,97 $\pm$ 0,03	7,5 $\pm$ 0,4	787 $\pm$ 5	302 $\pm$ 4	1096 $\pm$ 8	0,38 $\pm$ 0,00
<i>ccoamtCp 14-4</i>	14,32 $\pm$ 0,03	6 $\pm$ 0,8	754 $\pm$ 44	362 $\pm$ 30	1121 $\pm$ 75	0,48 $\pm$ 0,01

partir des lignines d'Arabidopsis est systématiquement un peu plus élevée dans le mutant *ccoamt-1* et dans les lignées complémentées. Cette augmentation de la fréquence de ces unités H minoritaires est en accord avec les résultats publiés sur le mutant *Atccomt-1* (Do, Pollet et al. 2007) et confirme que le défaut de biosynthèse des unités G entraîne une augmentation non seulement des unités S majeures, mais aussi des unités H mineures. Cependant, la principale différence concerne le rapport S/G, de ces lignées qui atteint des valeurs supérieures à la fois au témoin Col0 et à la lignée issue de l'introduction de la construction dans ce témoin. Dans leur ensemble, les résultats obtenus sur transformants T1 (maintien du rapport S/G et du % de H plus élevés) ne supportent pas l'hypothèse d'une complémentation fonctionnelle du mutant par le gène *BdCCoAOMT1*.

Trois lignées descendant des lignées analysées précédemment (la troisième génération, T3) ne contenant qu'une seule insertion de *BdCCoAOMT1* à l'état homozygote ont été obtenues. L'analyse par Klason de la quantité globale de lignine montre que celle-ci n'est pas rétablie dans les lignées T3 et reste à un niveau similaire à celui du mutant, indiquant l'absence de complémentation de la lignée *Atccomt-1* par *BdCCoAOMT1*.

## 6. Identification de mutants *Bdccoamt1* par TILLING

Le gène *BdCCoAOMT1* a été défini comme gène cible du TILLING, comme décrit au chapitre III. Les résultats sont présentés Tableau 4.

Deux familles portant des mutations faux-sens dans la séquence codante de *BdCCoAOMT1* ont été identifiées : 3631 et 5660. Des mutations ont été détectées dans quatre autres familles mais n'ont pas pu être caractérisées directement. Elles le seront lors du génotypage de ces familles.

**Tableau 4.** Récapitulatif et description des mutations dans le gène *BdCCoAOMT* et leurs conséquences sur la séquence protéique dans les familles identifiées par TILLING.

Position sur l'ADNc	Position sur la protéine	Type de mutation	Famille	Propriétés physiques AA sauvage	Propriétés physiques AA muté	Scores SIFT
G157A	V53M	<b>faux-sens</b>	<b>3631</b>	Apolaire, hydrophobe, aliphatique, petit	Apolaire, hydrophobe	0.08
C192T	L64L	L64L	5660*	Apolaire, hydrophobe, aliphatique		
C193A	R65S	<b>faux-sens</b>	<b>5660*</b>	Polaire, basique (chargé +)	Minuscule, polaire	0.05
>650bp/440bp			3692			
G241A	D81N	<b>faux-sens</b>	<b>3590</b>	Petit, acide (chargé -)	Petit, polaire	0.07
538bp/ ?			3587			
469bp/480bp			4024			

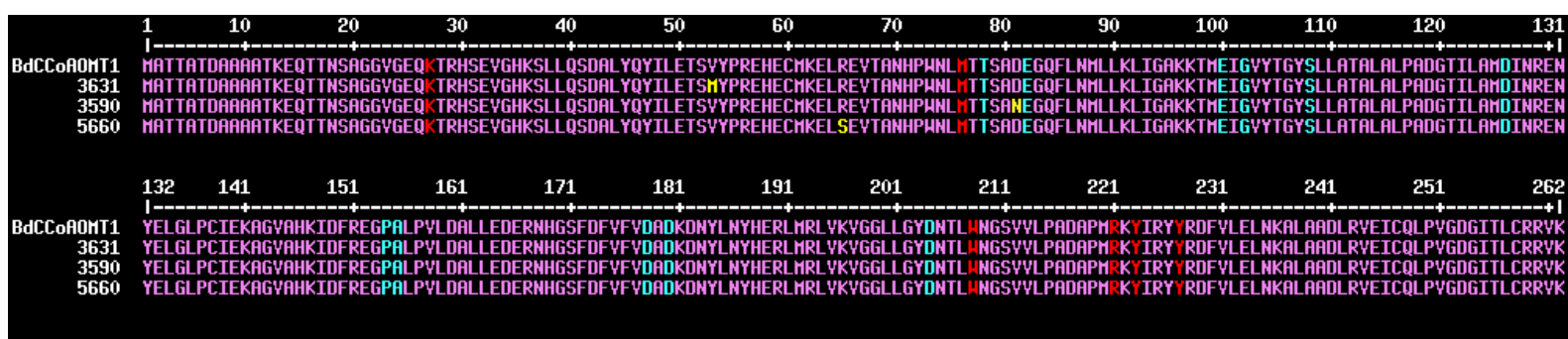
**Tableau 5.** Etude des lignines des tiges matures des lignées homozygotes mutantes 5660 (moyenne de 7 plantes), 3590 (moyenne de 4 plantes) comparées au sauvage Bd21-3 (moyenne de 3 plantes). Les analyses sont réalisées sur les résidus pariétaux obtenus à partir des tiges récoltées à maturité. Le taux de lignine Klason (LK) est exprimé en % pondéral du RP. Le rendement total en monomères de thioacidolyse est exprimé en  $\mu$ moles par gramme de LK. Les quantités d'acide p-coumarique (PC) et d'acide férulique (Fe) libérés par thioacidolyse sont exprimées en mg/g de RP. Les données ne sont pas significativement différentes entre le sauvage et les mutants.

	%LK en RP	Monomères H, G et S de thioacidolyse					Acides p-OH cinnamiques (en mg/g RP)	
		Total ( $\mu$ mole/g LK)	S/G	%H	%G	%S	PC	Fe
Bd21-3	17,02 $\pm$ 0,16	1584 $\pm$ 124	2,38 $\pm$ 0,17	3,6 $\pm$ 0,0	28,6 $\pm$ 1,4	67,8 $\pm$ 1,4	9,6 $\pm$ 1,6	9,6 $\pm$ 1,3
5660	16,39 $\pm$ 0,48	1366 $\pm$ 77	2,18 $\pm$ 0,10	3,8 $\pm$ 0,3	30,3 $\pm$ 1,0	65,9 $\pm$ 0,9	8,6 $\pm$ 0,8	9,7 $\pm$ 0,9
3590	16,78 $\pm$ 0,31	1478 $\pm$ 125	2,10 $\pm$ 0,13	3,6 $\pm$ 0,1	31,2 $\pm$ 1,3	65,2 $\pm$ 1,3	9,1 $\pm$ 0,5	10,6 $\pm$ 0,7

## 6.1. Prédiction de l'effet des mutations

- La mutation portée par la famille 3631 provoque la substitution d'une valine en méthionine en position 53 de la séquence protéique. Le score SIFT de cette substitution est de 0.08. Cette valeur étant supérieure à 0.05, elle est considérée comme tolérante, c'est-à-dire qu'elle n'a pas d'effet délétère prédit sur l'activité de la protéine. Sur la séquence primaire, cette substitution est relativement éloignée de l'acide aminé essentiel le plus proche (à 17 acides aminés de la Lysine en position 36). Sur la prédiction de structure tridimensionnelle, elle se situe en face d'une hélice portant un résidu impliqué dans la fixation du cofacteur SAM, mais celui-ci est situé de l'autre côté de l'hélice montrée par la visualisation en modèle « cartoon » présentant les éléments de structure secondaire de la protéine (Fig.16a). La visualisation en plus gros plan et en modèle « sphère » où l'espace occupé par un résidu est représenté sous forme de sphères (Fig. 16b) suggère que l'encombrement causé par la différence de taille pourrait modifier légèrement la position de l'hélice et par conséquent l'organisation du site de fixation au cofacteur. Cependant, un tel effet est peu probable. La substitution est également située à proximité du site de dimérisation mais la chaîne latérale du résidu concerné est orientée vers l'intérieur de la protéine (pas en direction du second monomère) et est déjà probablement trop éloignée pour avoir un effet.

- La mutation touchant la famille 5660 provoque la substitution en position 65 d'une arginine en sérine. Son score SIFT est de 0.05, ce qui est la valeur limite de détermination de tolérance d'une substitution. Cet outil ne prédit donc pas d'effet de cette substitution sur l'activité de BdCCoAOMT1. Cette substitution est située face à une boucle portant un résidu impliqué dans la fixation du cofacteur et un résidu impliqué dans la reconnaissance du substrat (Fig. 16 c). La poche de fixation du cofacteur et du substrat se trouvant de l'autre côté de la boucle, pas au niveau du résidu



**Figure 6.** Séquence primaire de BdCCoAOMT1 et de ses allèles mutants identifiés par TILLING, sur laquelle les acides aminés essentiels à l'activité de la protéine ont été indiqués: en rouge, les résidus de fixation du substrat sur le site actif, en bleu clair les motifs de fixation au cofacteur et à l'ion Ca<sup>2+</sup>. Les acides aminés modifiés sont indiqués en jaune dans les lignes mutantes.

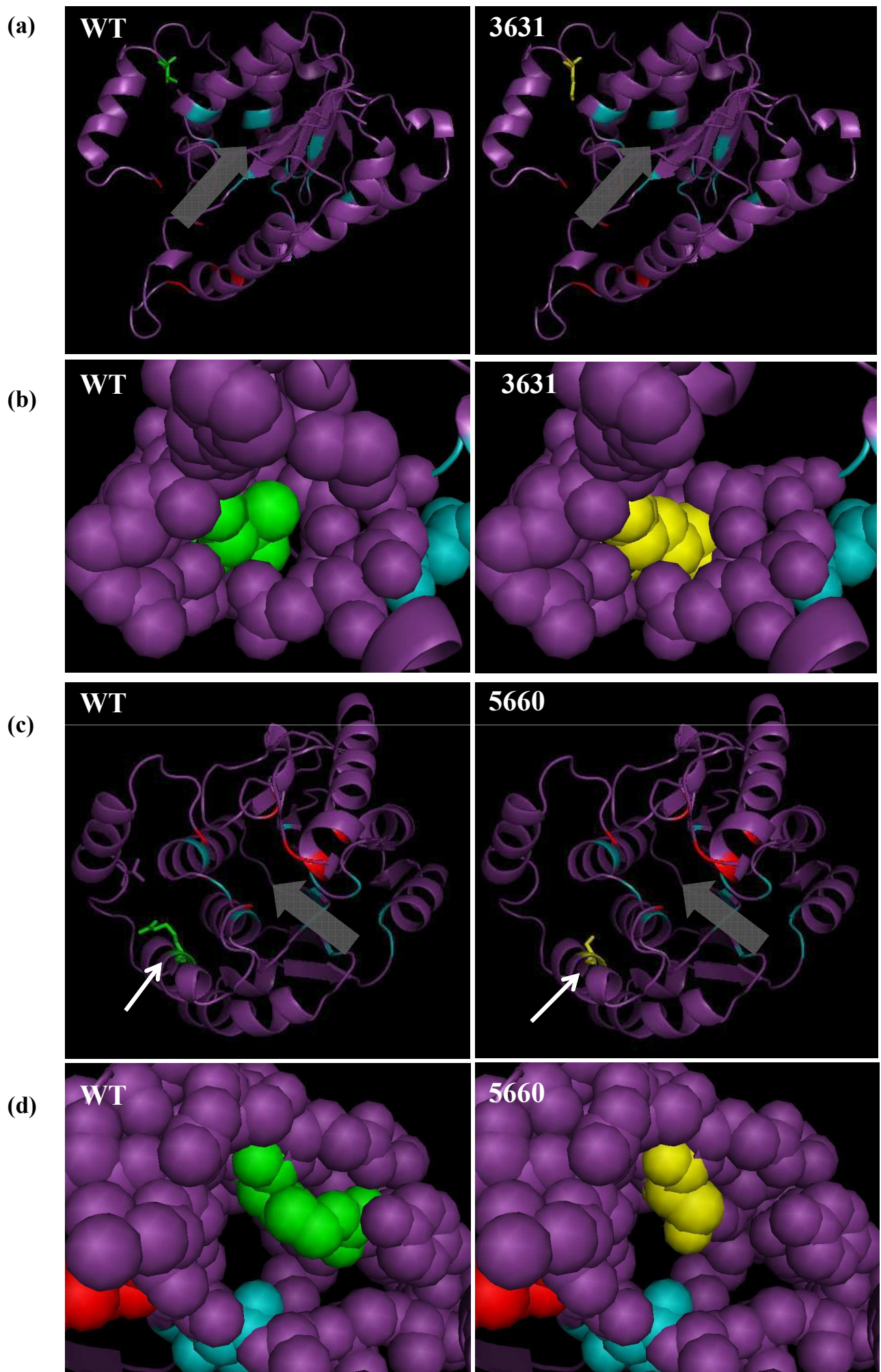
muté, un effet de cette substitution paraît peu probable. D'autant plus que la sérine est de plus petite taille que l'arginine et d'après la représentation en sphères ne semble pas modifier l'organisation des résidus l'entourant (Fig. 16d).

- La mutation touchant la famille 3590 substitue l'acide aspartique en une asparagine en position 81. Son score SIFT est de 0.07, ce qui est considéré comme tolérant, c'est-à-dire qu'elle n'aura vraisemblablement pas d'impact sur l'activité de la protéine. Elle est localisée juste à côté d'un résidu impliqué dans la fixation du cofacteur mais la modélisation 3D (Fig. 17a) montre qu'elle est située à l'extérieur de la poche de fixation du cofacteur et du substrat. La représentation en sphères (Fig. 17b) suggère que l'organisation des résidus entourant la substitution ne devrait pas être modifiée par l'encombrement dû à la présence de l'asparagine. De par sa nature et sa position, cette mutation ne devrait pas affecter BdCCoAOMT1.

## 6.2. Analyse des lignines des lignées identifiées

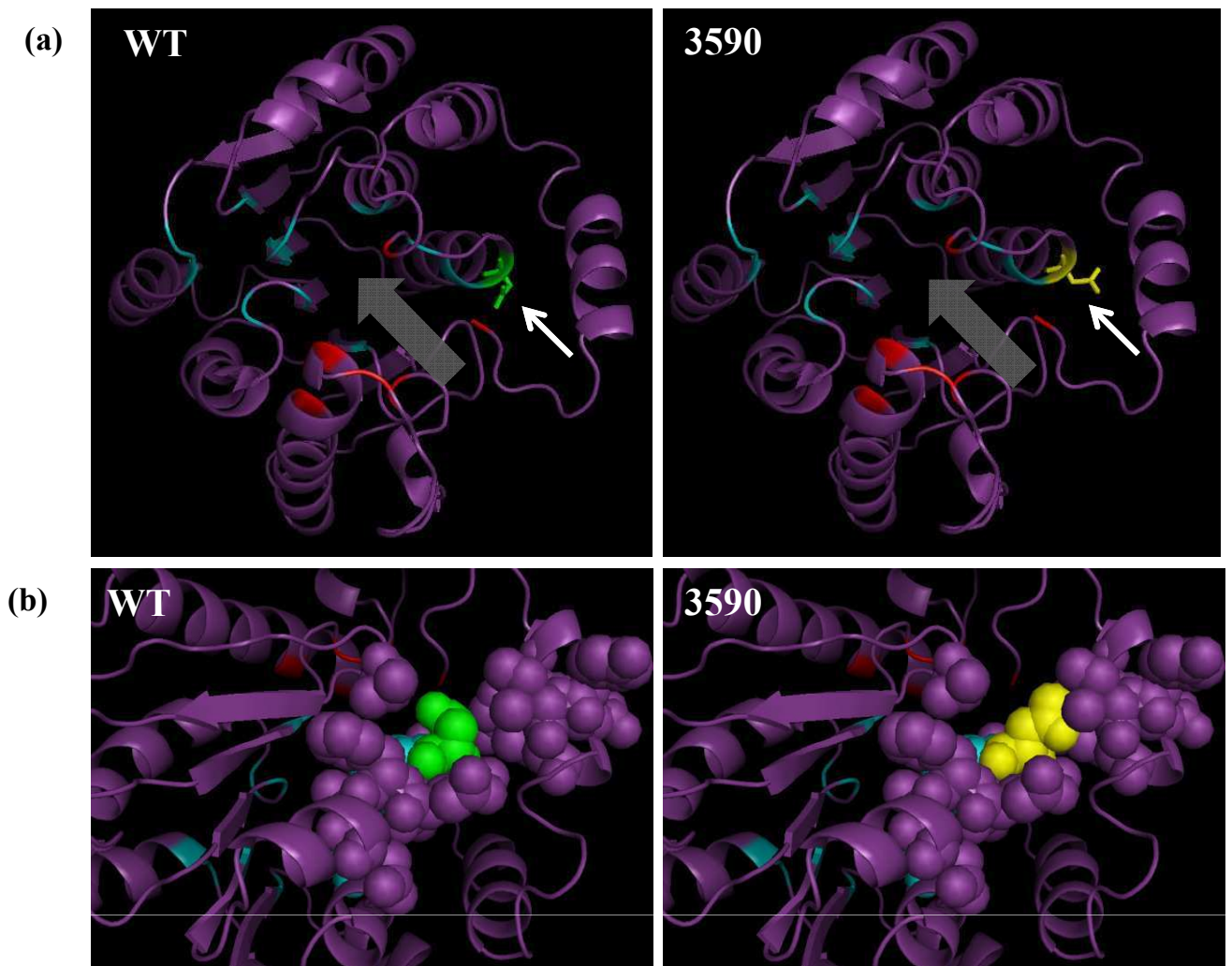
Des lignées homozygotes pour la mutation qu'elles portent dans le gène *BdCCoAOMT1* ont été identifiées pour deux familles jusqu'à présent : 5660 et 3590. L'étude des lignines a été réalisée sur les tiges matures débarrassées des composés solubles par extraction au solvant. Les taux de lignine Klason des résidus pariétaux ainsi obtenus sont assez similaires pour les deux mutants et pour le témoin Bd21-3 (Tableau 5). L'analyse de variance (ANOVA) suggère cependant que le taux du témoin est légèrement supérieur au taux de lignine Klason du mutant 5660 (95% confidence level), résultat à confirmer (le nombre de réplicats biologiques étant seulement de 3 pour le témoin et de 7 pour le mutant). L'étude de la structure des lignines par thioacidolyse ne révèle aucune différence significative en termes de rendement de thioacidolyse ou de rapport S/G. Il en est de même pour l'acide *p*-coumarique ou l'acide férulique liés aux parois et libérés par thioacidolyse (Tableau 5). Dans les tiges mûres et par rapport au témoin, les mutants





**Figure 7.** Prédiction de la structure tridimensionnelle de BdCCoAOMT1 chez Bd21-3 et chez les familles 3631 et 5660.





**Figure 8.** Prédiction de la structure tridimensionnelle de BdCCoAOMT1 chez le sauvage Bd21-3 et chez la famille 3590.

**Figure 7 et 8.** Prédiction de la structure tridimensionnelle de BdCCoAOMT1 sauvage et mutante visualisées dans PyMol. Les acides aminés d'importance particulière ont été représentés d'après le même code couleur que Fig. 6: en rouge les acides aminés de fixation du substrat sur le site actif, en bleu clair les motifs de fixation au cofacteur et au  $\text{Ca}^{2+}$ . Les acides aminés modifiés sont indiqués en jaune dans les lignées mutantes et en vert sur la séquence sauvage. Les protéines ont été orientées de façon à montrer la position du résidu muté dans chaque famille, indiqué par une flèche blanche. Les flèche pleines grises indiquent la poche de fixation su substrat et du cofacteur.

La représentation en cartoon (a) et (c) montre l'assemblage des structures secondaires (hélices  $\alpha$  et feuillets  $\beta$ ) mais ne rend pas compte de l'encombrement des résidus. La représentation en sphères (b) et (d) permet de visualiser en gros plan l'organisation dans l'espace des résidus mutés et leurs voisins.

étudiés ne présentent donc aucune altération globale nette en terme de lignines ou en terme de dérivés de la lignification de leurs tiges ou des dérivés *p*-hydroxycinamiques liés aux parois.

## 7. Conclusion et perspectives

Le gène *BdCCoAOMT1* (*Bradi1g48370.2*) nous semblait être le meilleur candidat pour la voie de biosynthèse des monolignols d'après les comparaisons de séquences et l'analyse phylogénétique où il était le seul gène de *Brachypodium* regroupé avec d'autres gènes *CCoAOMT*. L'analyse des profils d'expression de plusieurs autres candidats *CCoAOMT* a révélé que *BdCCoAOMT3* (*Bradi3g39380.1*) serait un candidat plus probable. Cependant, *BdCCoAOMT1* pourrait avoir une activité 5-hydroxyferuloyl-coenzyme A 3-*O*-méthyltransférase qui reste à démontrer. La suite des travaux de TILLING portera donc sur la recherche de mutants du gène *BdCCoAOMT3*, en raison d'une part de l'impact de l'activité *CCoAOMT* lignine-spécifique sur le taux de lignines (plus faible) et sur la structure de ces polymères (rapport S/G plus élevé) établi chez les dicotylédones (Guo, Chen et al. 2001; Pinçon, Maury et al. 2001; Do, Pollet et al. 2007), et, d'autre part, de son impact potentiel sur les esters féruliques qui réticulent les parois de graminées (aspect plus spéculatif).

Il a été montré dans la partie précédente que les mutations portées par les familles analysées avaient une faible probabilité de modifier l'activité de *BdCCoAOMT1*. Pour l'analyse de plusieurs lignées portant des mutations ciblant les sites essentiels à l'activité de la protéine ou générant un codon STOP précoce, empêchant ainsi toute activité de la protéine permettrait de résoudre les incertitudes sur la fonction biologique de ce gène. Ce type de mutation a déjà été observé dans d'autres gènes dans plusieurs familles au sein de la population, et en criblant un nombre suffisant de familles, il est probable d'en trouver. Malgré cela, si ce gène est peu exprimé et n'a qu'une participation annexe à la voie de biosynthèse des lignines des tiges, les

**Tableau 6.** Localisation sur la séquence de BdCCoAOMT1 et description des acides aminés trouvés spécifiquement sur les séquences des CCoAOMT de monocotylédones utilisées pour les alignements (Fig. 1 et 2) et l'étude phylogénétique (Fig. 3) décrits dans ce chapitre, comparés aux CCoAOMT des autres espèces utilisées.

Position sur BdCCoAOMT1	AA monocot	propriétés	AA autres espèces	propriétés
60	C ou S	Petits et polaires	A ou P	Petits et apolaires
94	G	Apolaire, aliphatique	N	polaire
96	K	Polaire et basique	N	polaire
121	T	polaire	K	Polaire et basique
138	C ou S	Petits et polaires	I ou V	Apolaires aliphatiques
189	E	Polaire et acide	K	Polaire et basique
193	R ou K	Basiques	D ou E	Acides

modifications de lignines dans les mutants pourront être très subtiles et n'auront potentiellement aucun effet sur les rendements de saccharification de leurs parois.

Cependant, cette partie de l'étude illustre les limites de l'annotation de gène par homologie et analyse phylogénétique. L'observation de la figure 2 suggérait la présence de résidus spécifiques aux quatre CCoAOMT de graminées alignées. Ceux-ci sont récapitulés Tableau 6. Les différences importantes de propriétés physiques (systématiquement au niveau de la polarité ou de la charge) de ces acides aminés, qui semblait ne représenter qu'une spécificité des graminées par rapport aux autres plantes alignées (dicotylédones et une gymnosperme), pourrait en fait être le signe d'une activité ou une spécificité de substrat différentes (la forte conservation de la majorité des résidus et les similitudes au sein des trois séquences de graminées écartent à priori l'hypothèse d'une perte totale de fonction). En effet, les CCoAOMT de riz et de maïs utilisées dans cette étude n'ont été annotées que sur la base de leur séquence et de leur expression (Civardi, Rigau et al. 1999; Tanaka 2008), pas sur leur activité. Ces observations semblent indiquer que l'identification des *CCoAOMT* chez les graminées est à réviser, et pour ce faire, *BdCCoAOMT3* serait un point de départ possible.

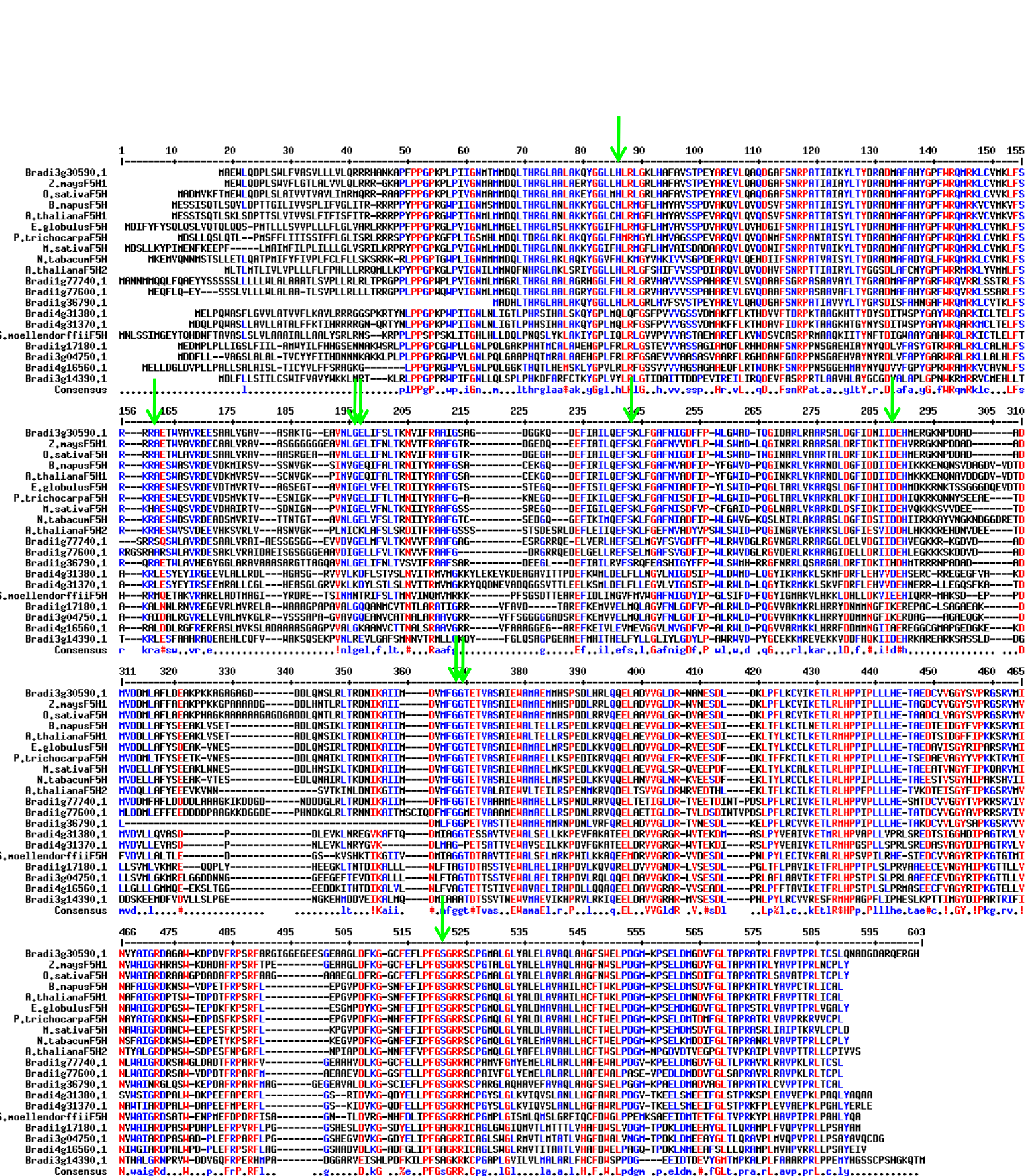


## Chapitre V :

### Identification de la férulate-5-hydroxylase impliquée dans la voie de biosynthèse des monolignols

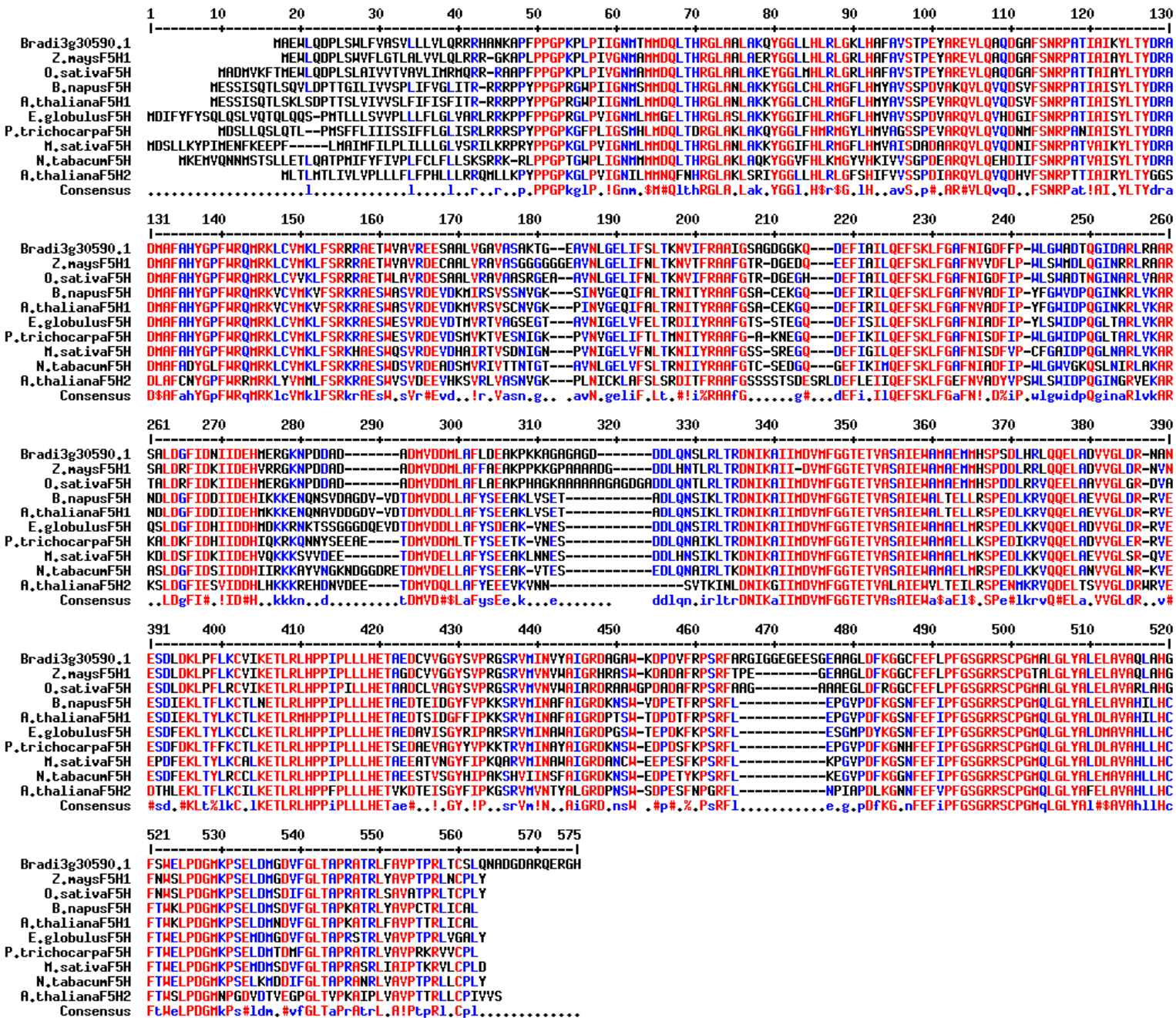
## Chapitre V : Identification de la férulate-5-hydroxylase impliquée dans la voie de biosynthèse des monolignols

Ce chapitre présente la détermination du gène candidat pour la férulate 5 hydroxylase (F5H) impliquée dans la voie de biosynthèse des monolignols.



**Figure 1.** Alignement des candidats F5H de *Brachyodium* avec ceux de F5H d'autres espèces végétales. Les flèches vertes indiquent des de résidus très conservés chez les F5H de ces espèces mais qui ne sont pas retrouvés dans les séquences de *Brachyodium* à l'exception de Bradi3g30590.1. En rouge: résidus à fort consensus; en bleu: résidus à consensus moyen, en noir les résidus neutres.





**Figure 2.** Alignment of candidate F5H Bradi3g30590.1 avec des F5H impliquées dans la lignification chez d'autres espèces végétales. La couleur rouge montrant les forts consensus, on distingue la grande similarité de séquence entre ces protéines. Sont indiqués en rouge les résidus à fort consensus; en bleu les résidus à consensus moyen et en noir les résidus neutres.

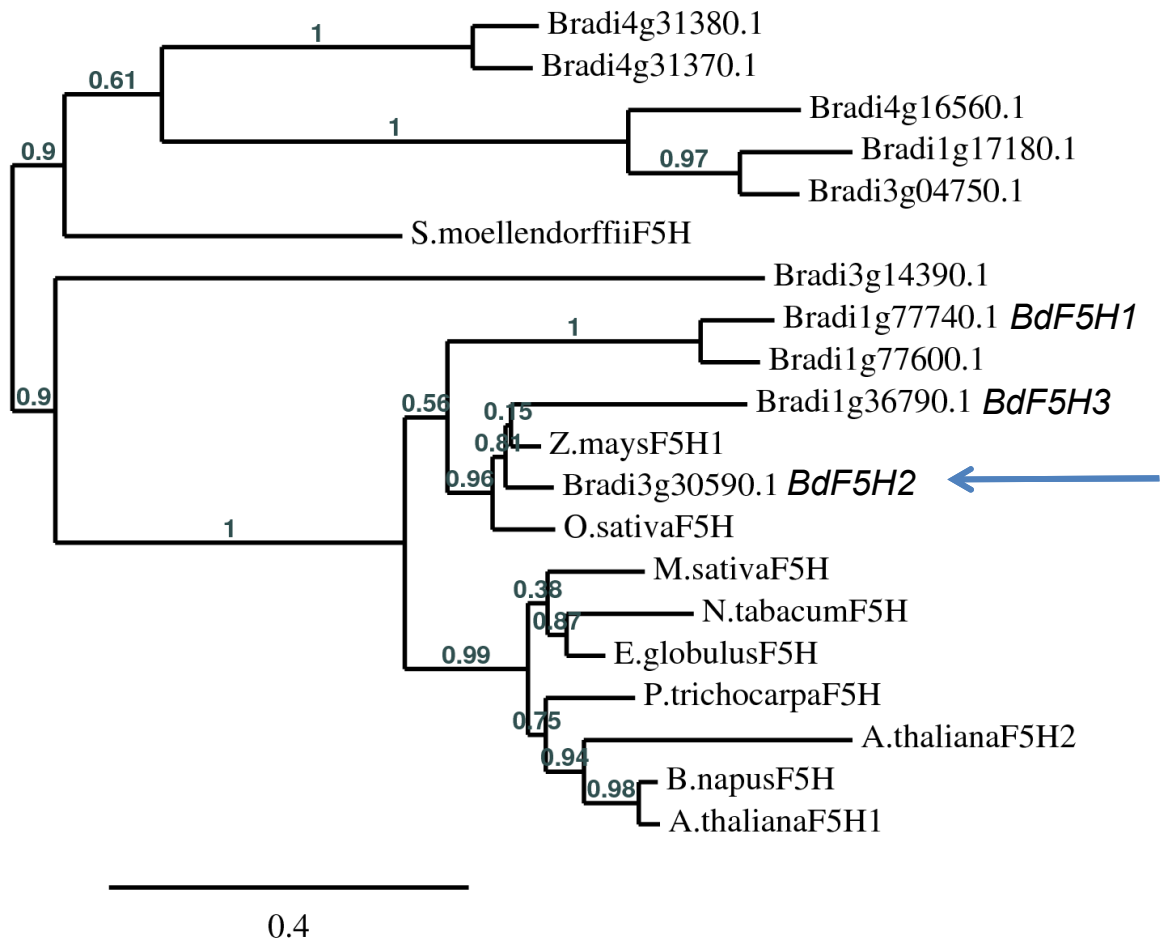
La férulate-5-hydroxylase est la mono-oxygénase cytochrome P450-dépendante qui dirige la voie de biosynthèse de monolignols vers la formation de l'alcool sinapylique en catalysant la troisième étape d'hydroxylation du noyau aromatique.

### 1. Détermination du gène candidat par alignements de séquences

La protéine codée par le gène *AtF5HI* (locus : *At4g36220*) a été utilisée comme séquence d'entrée au BLAST sur le génome 4X, puis 8X de *Brachypodium distachyon*. Parmi les gènes de *Brachypodium* issus de cette recherche, ceux possédant les scores les plus élevés et les e-value les plus faibles sont les suivants :

	Score	P(N)
Bradi3g30590.1	1739	3.4e-180
Bradi1g77740.1	1403	1.4e-144
Bradi1g36790.1	666	7.0e-128
Bradi1g77600.1	960	1.2e-97
Bradi4g31380.1	909	3.1e-92
Bradi1g17180.1	866	1.1e-87
Bradi3g14390.1	842	3.8e-85

Les scores très élevés des deux premiers résultats suggèrent que les protéines pour lesquelles ils codent sont toutes les deux très similaires à *AtF5HI*. L'alignement de la séquence de ces 2 protéines sur celles d'autres espèces est présenté en Figure 1. *Bradi1g36790.1* peut être éliminé d'emblée des candidats en raison de sa taille : la protéine codée par ce gène est bien plus courte que les autres et il lui manque un certain nombre d'acides aminés en début et milieu de séquence, acides aminés par ailleurs très conservés dans toutes les espèces comparées. Cette protéine n'est probablement pas fonctionnelle et sa présence en troisième position du résultat du BLAST vient de la très faible P-value (P(N)) : cette valeur représente la probabilité que la similitude des séquences soit due au hasard. Un certain nombre de résidus qui apparaissent très conservés dans les protéines alignées ne sont pas retrouvés dans les séquences de celles codées par les gènes de *Brachypodium*, à part dans le cas de *Bradi3g30590.1* qui devient alors le choix



**Figure 3.** Arbre phylogénétique construit à partir des séquences protéiques de F5H identifiées dans plusieurs espèces, incluant celles codées par les gènes candidats de *Brachypodium distachyon*. Les chiffres gris indiquent la valeur « bootstrap » représentant la probabilité d’existence de chaque branche du phylogramme lors des répétitions des calculs de construction. La longueur des branches est proportionnelle à la distance évolutive entre les nœuds. La flèche bleue indique le gène candidat de *Brachypodium* retenu dans cette étude.

**Tableau 1.** Correspondance entre les différentes appellations des gènes *F5H* annotés chez *Brachypodium distachyon*, accompagné du nombre d’EST correspondant à chacun de ces gènes.

Nom d’après leur position sur le génome	Nomenclature simplifiée	EST
Bradi1g77740.1	BdF5H1	0
Bradi3g30590.1	BdF5H2	1
Bradi1g36790.1	BdF5H3	0

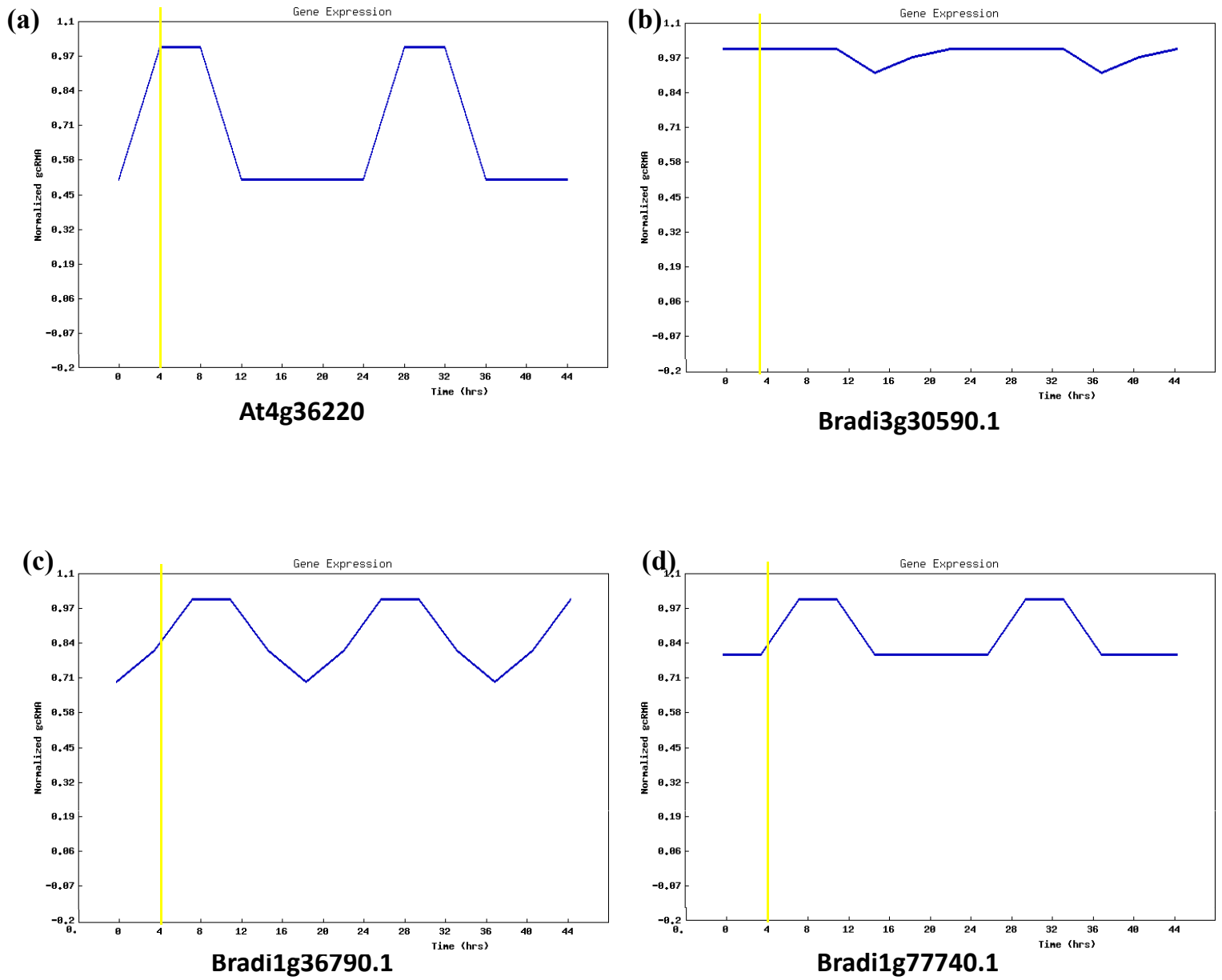
initial de candidat de *F5H* chez *Brachypodium*. Quelques exemples ayant poussé à ce constat sont indiqués par une flèche verte sur la Fig. 1.

La Figure 2 présente l'alignement de la séquence de la protéine codée par Bradi3g30590.1 sur celles de F5H impliquées dans la lignification d'autres espèces végétales. Cette analyse confirme la forte similarité de séquence entre cette protéine et les F5H spécifiques de la lignification d'autres espèces végétales, similarité soulignée par la couleur rouge sur l'alignement de la Fig. 2.

## 2. Etude phylogénétique

L'intégralité des séquences protéiques alignées dans le paragraphe précédent a été soumise à une analyse phylogénétique dont le résultat est donné en Fig. 3 sous forme de phylogramme. On observe qu'à l'exception de celle de *Selaginella moellendorffii* qui a évolué indépendamment des F5H des angiospermes (Weng, Banks et al. 2008), les séquences des F5H comparées sont regroupées dans un cluster au sein duquel se distinguent celles appartenant aux dicotylédones de celles appartenant aux graminées. Dans ce dernier groupe, deux séquences de *Brachypodium* sont présentes : Bradi3g30590.1 qui était le candidat le plus vraisemblable d'après les alignements de séquence et Bradi1g36790.1 qui avait été éliminé à cause de sa taille. Malgré sa présence au sein du groupe, cette éviction nous semble d'autant plus légitime que cette F5H est la plus éloignée de la F5H de maïs et de riz, éloignement illustré par la longueur de la branche la portant (la longueur des branches du phylogramme est proportionnelle aux différences évolutives entre les séquences).

Trois gènes *F5H* ont été officiellement annotés sur la séquence génomique de *Brachypodium* : *Bradi1g36790.1*, *Bradi1g77740.1* et *Bradi3g30590.1* qui seront respectivement appelés *BdF5H1*, *BdF5H2* et *BdF5H3* (Tableau 1). La recherche d'EST (Expressed Sequence Tag, court transcrit montrant l'abondance de l'expression du gène lui correspondant) provenant de



**Figure 4.** Expression en fonction du cycle circadien des gènes *AtF5H1* (a), *BdF5H2* (b), *BdF5H3* (c) et *BdF5H1* (d) dans les conditions de culture les plus proches de celles de Versailles. Le temps en heures est indiqué en abscisse, et une unité arbitraire normalisée est indiquée en ordonnée. Les barres jaunes indiquent quatre heures après le début de l'exposition à la lumière.

l'expression de ces gènes a renforcé le choix de *BdF5H2* comme candidat : il en présente un alors que *BdF5H1* et *BdF5H3* n'en ont aucun (Tableau 1).

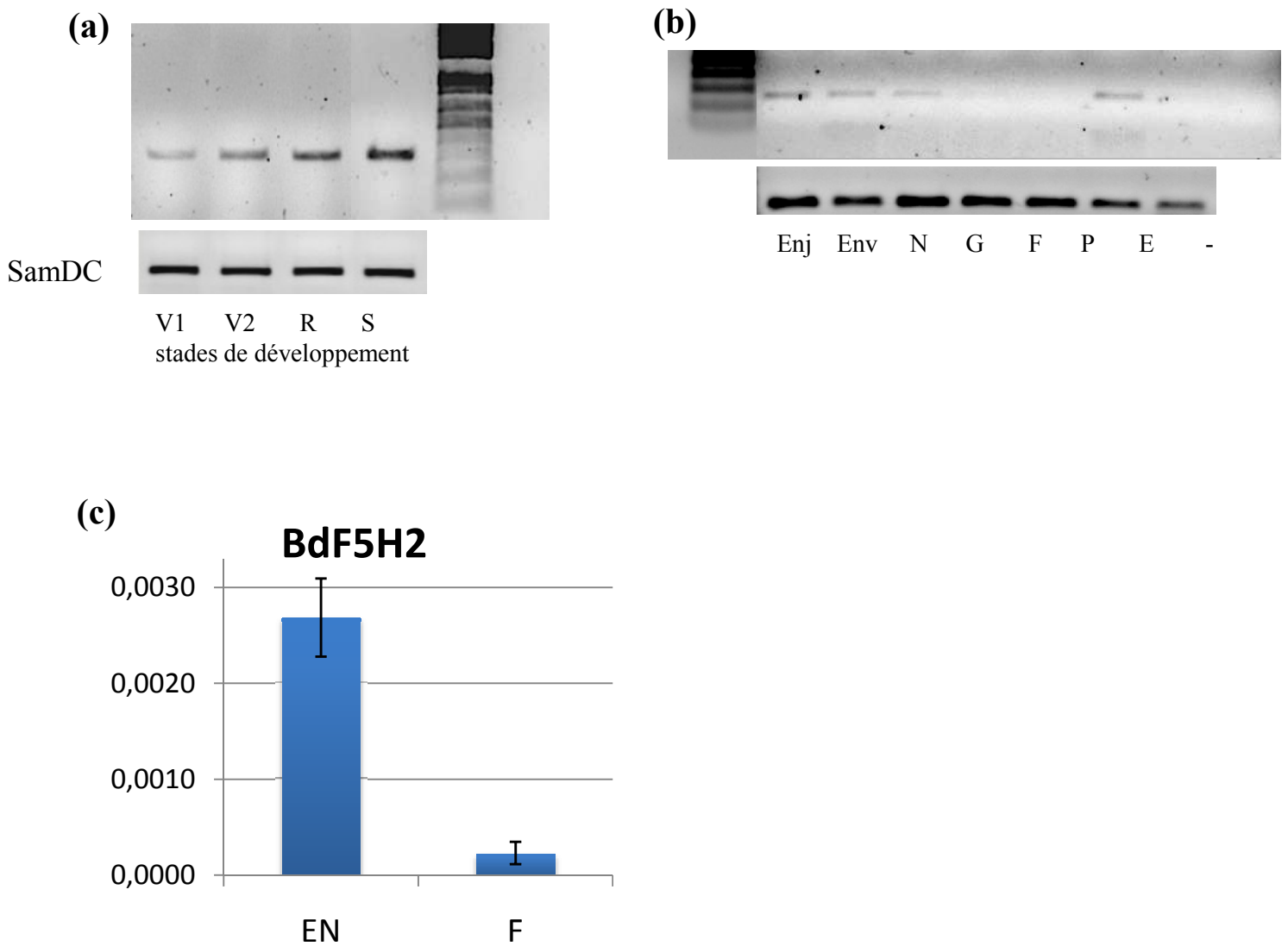
### 3. Expression en fonction du cycle circadien

Le gène de *F5H* de riz n'étant pas présent dans les données du serveur Diurnal, *AtF5H1* a été choisi comme base de comparaison. Comme les autres gènes de la voie de biosynthèse des monolignols, il présente un plateau d'expression au bout de quatre heures d'éclairage. Le profil de *BdF5H2* est très différent, étant déjà à sa valeur maximale à l'éclairage, celle-ci chutant 12 heures plus tard. Quatre heures après éclairage, les deux autres gènes présentent des valeurs minimales ou intermédiaires de leurs niveaux d'expression. Bien que décalé, le profil ressemblant le plus à celui d'*AtF5H1* serait celui de *BdF5H1*. Cependant, cet argument n'est pas suffisant pour modifier le candidat choisi.

### 4. Profils d'expression

#### 4.1. RT-PCR

Les ARN de tiges entières (nœuds, entre-nœuds et gaines foliaires) ont été extraits à quatre stades de croissance : deux stades végétatifs (V1 et V2, Fig. 5), au début du stade reproductif (R, Fig. 5) et peu avant le début de sénescence (S, Fig. 5). Après rétro-transcription, ils ont été amplifiés par RT-PCR semi-quantitative avec des oligonucléotides spécifiques du gène *BdF5H2* (Fig 5a). Les résultats montrent une augmentation de l'expression de *BdF5H2* au fil de ces stades, ce qui est cohérent avec le profil d'un gène « lignine ». Les ARN d'organes distincts ont ensuite été extraits au stade R, rétro-transcrits puis amplifiés avec les mêmes oligonucléotides (Fig. 5b). Cette analyse suggère que l'expression de ce gène est la plus forte dans les entre-nœuds jeunes en cours de croissance, et dans le pédoncule, puis par ordre



**Figure 5.** Etudes d'expression du gène *BdF5H2* chez *Bd21-3* (a) RT-PCR semi quantitative sur les tiges de *Bd21-3* aux quatre stades suivants: fin du stade végétatif à 5 semaines (V1), avant émergence de l'épi à 7 semaines (V2), après émergence de l'épi à 9 semaines (R) et avant sénescence à 12 semaines (S), ciblant spécifiquement *BdF5H2*. Les trois premiers stades apparaissent similaires, avec une expression qui semble un peu plus intense à 9 semaines. Une seule des 3 répétitions biologiques effectuées est montrée. (b) RT-PCR au stade 7 semaines sur différents organes du maître brin de trois plantes. Enj: jeunes entre-nœuds (en cours de lignification); Env: entre-nœuds lignifiés; N: nœuds; G: gaine foliaire; F: feuilles; P: péduncule; E: épi. (c) RT-PCR quantitative montrant les niveaux d'amplification de *BdF5H2* rapportés à ceux du gène de ménage *SamDC* dans les entre-nœuds (EN) et les feuilles (F) de quatre répétitions biologiques de *Bd21-3* après émergence de l'épi. (a) (b) L'amplification du gène constitutif *SamDC* permet d'indiquer la quantité de matrice présente dans chaque puits. (c) Les données ont été normalisées par rapport au gène constitutif *SamDC*.

décroissant, dans les entre-nœuds âgés (déjà lignifiés), les nœuds, et dans la gaine foliaire. En revanche, son expression n'est pas détectable dans les feuilles et les épis. Ces résultats sont appuyés par RT-PCR quantitative (Fig. 5C) montrant que l'expression dans les feuilles est environ dix fois moins forte que dans les entre-nœuds. Ces résultats sont cohérents avec un gène impliqué dans la lignification, et confirment le choix de *BdF5H2* comme candidat.

#### 4.2. Bradinet

Les données d'expression obtenues sur Bradinet (Fig. 6) indiquent que l'expression du gène *BdF5H2* est au moins dix fois supérieure à celle des deux autres *F5H* annotées. En accord avec les résultats du paragraphe précédent, il s'exprime majoritairement dans les entre-nœuds, le pédoncule et le mésocotyle. Ces données valident le choix de *BdF5H2* comme candidat.

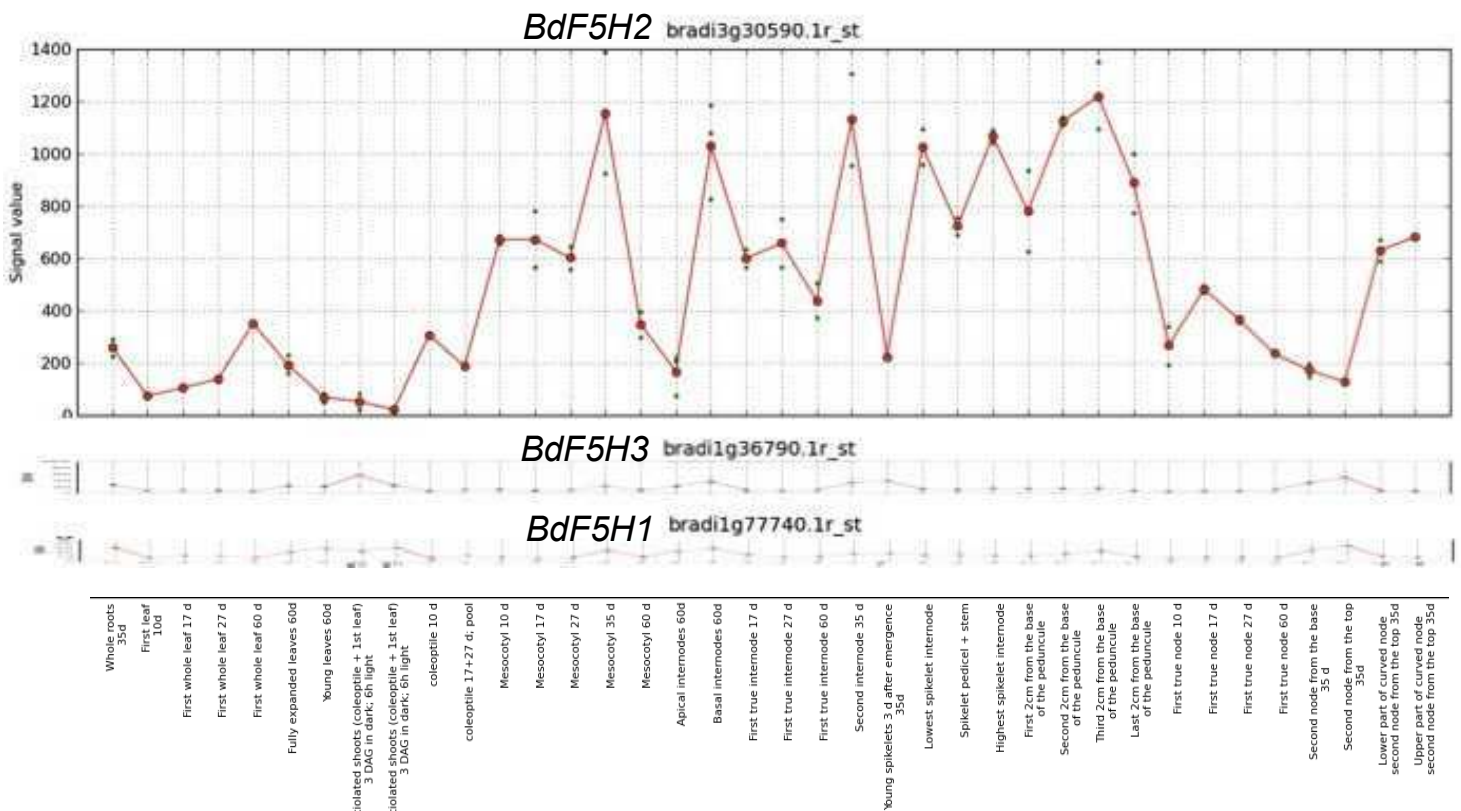
### 5. Complémentation fonctionnelle

Afin de déterminer si la protéine codée par le gène *BdF5H2* possède une activité équivalente à celle de la férulate-5-hydroxylase *F5H1* d'*Arabidopsis thaliana*, il a été cloné sous contrôle du promoteur constitutif 35S du CaMV puis introduit dans la lignée *Atf5h1*(EBL87).

#### 5.1. Analyses des lignines des tiges matures

La perte de fonction du gène *F5H1* chez le mutant EBL87 d'*Arabidopsis thaliana* conduit, d'après les résultats de thioacidolyse, à la disparition quasi-totale des unités S (rapport S/G très proche de zéro et fréquence des monomères S de thioacidolyse de l'ordre de 0,1%) (Tableau 2). Quatre lignées transformées ont été sélectionnées et l'analyse des lignines par thioacidolyse révèle une augmentation du rapport S/G, rapport restauré au niveau de celui du sauvage dans la lignée *f5hcp20*, et bien supérieur à celui du sauvage dans les autres lignées. Le rapport S/G de la lignée *f5hcp6* atteint 7.63. Ce chiffre remarquablement élevé correspond à une fréquence





**Figure 6.** Niveaux d'expression des trois *F5H* annotées de *Brachypodium* dans des organes et à des stades différents, provenant de Bradinet. Afin de faciliter les comparaisons, les trois graphes sont à la même échelle en ordonné (échelle indiquant 500 unités d'expression). Les différents tissus sont indiqués en abscisse.

d'unités S proche de 90% dans la fraction non condensée des lignines (fraction qui génère les monomères de thioacidolyse) tandis que cette fréquence est voisine de 30% dans le cas de WS. La thioacidolyse révèle en outre que les monomères 5-OH G sont obtenus en quantité plus élevée à partir des lignées complémentées dont le rapport S/G est très élevé qu'à partir de WS. Ces résultats suggèrent que l'expression du transgène sous contrôle d'un promoteur fort génère une activité F5H à un niveau anormalement élevé. Le 5-hydroxyconiféraldéhyde issu de cette activité est alors formé en quantité trop importante pour être entièrement pris en charge par la COMT, raison pour laquelle on voit apparaître une teneur inhabituelle en unités 5-OH G. Ces résultats sont en accord avec des résultats obtenus sur des lignées d'*Arabidopsis* surexprimant la *F5H* et chez lesquelles une teneur anormalement élevée en unités 5-OH G avait été mise en évidence par RMN (Marita, Ralph et al. 1999). Les lignées complémentées présentent un léger défaut de croissance et sont légèrement plus petites que le contrôle WS. Ce trait est particulièrement accentué pour la lignée *f5hcp6* présentant le rapport S/G le plus élevé. Pour deux de ces lignées, des descendants ne contenant qu'une seule insertion du transgène à l'état homozygote ont été sélectionnés (génération T3). Les analyses de lignines sur ces lignées, descendant de *f5hcp6* et *f5hcp15*, confirment l'augmentation importante du rapport S/G observée à la première génération, ainsi que la fréquence accrue des unités 5-OH G (Tableau 3). Ce bouleversement de la structure des lignines est positivement corrélé à un enrichissement en unités liées seulement par des liaisons  $\beta$ -O-4, ainsi que le révèle l'augmentation substantielle du rendement de thioacidolyse, lorsque ce rendement est rapporté à la teneur en lignine des échantillons. Il est également accompagné d'une baisse substantielle de la teneur en lignines des tiges matures, évaluée par le dosage de Klason. Cette observation rejoint des résultats récents obtenus sur peuplier et montrant une corrélation négative entre le taux de lignines et la fréquence des unités S (Bose, Francis et al. 2009).

**Tableau 2.** Analyse comparée des lignines de la première génération de transformants (T1) correspondants à des lignées *d'Arabidopsis thaliana* mutantes *f5h* exprimant le gène *BdF5H2*, par rapport au témoin WS, au témoin WS ayant reçu la construction *BdF5h2* et au mutant *EBL87*. L'étude des lignines est réalisée en simple essai par thioacidolyse des tiges matures, avec l'objectif de cribler les lignées d'intérêt. Le rendement en monomères *p*-hydroxyphényles (H), guaiacyles (G) et syringyles (S) est exprimé en micromoles par gramme d'échantillon sec. En outre, le rendement en monomères 5-hydroxyguaiacyles (5-OH G) est indiqué.

	Rendement en $\mu\text{mol/g}$				S/G	% molaire			$\mu\text{mol 5OH-G}$ en $\mu\text{mol/g}$
	H	G	S	total		%H	%G	%S	
WS	1.1	103	46	151	0.45	0.7	68.6	30.7	0.6
<i>f5h</i> (EBL87)	0.6	116	0	116	0.00	0.5	99.3	0.1	0.1
WS+BdF5H2	1.7	28	112	141	4.07	1.2	19.5	79.3	0.3
<i>f5hcp</i> 6 ( <i>f5h</i> +BdF5H2)	1.7	18	135	154	<b>7.63</b>	1.1	11.5	87.4	4.1
<i>f5hcp</i> 9 ( <i>f5h</i> +BdF5H2)	0.9	24	63	88	2.66	1.0	27.1	71.9	1.3
<i>f5hcp</i> 15 ( <i>f5h</i> +BdF5H2)	0.9	44	73	118	1.66	0.7	37.2	62.0	1.4
<i>f5hcp</i> 20 ( <i>f5h</i> +BdF5H2)	0.6	83	35	118	0.42	0.5	70.3	29.2	0.5

**Tableau 3.** Etude des lignines de la troisième génération (T3) de lignées *d'Arabidopsis thaliana* *EBL87* complémentées (*f5hcp*). Les analyses sont réalisées sur les résidus pariétaux (RP) obtenus par extraction des tiges récoltées à maturité et les résultats correspondent aux valeurs moyennes et aux écarts moyens entre doubles biologiques. Le taux de lignine Klason (%LK) est exprimé en % pondéral du résidu pariétal (échantillon débarrassé de ses composés solubles). Le rendement total en monomères de thioacidolyse est exprimé en  $\mu\text{moles}$  par gramme de LK.

	% paroi	%LK	Rendement total	%H	%G	%S	%5-OH G	S/G
WS	56,8 $\pm$ 1,4	20,55 $\pm$ 0,21	1279 $\pm$ 14	0,4 $\pm$ 0,0	73,4 $\pm$ 1	26,0 $\pm$ 0,8	0,2 $\pm$ 0,0	0,35 $\pm$ 0,01
EBL87	61,7 $\pm$ 1,5	21,25 $\pm$ 0,11	1022 $\pm$ 2	0,4 $\pm$ 0,0	99,4 $\pm$ 0,1	0,2 $\pm$ 0,1	Tr	0,00 $\pm$ 0,00
<i>f5hcp</i> 6-11	54,85 $\pm$ 0,35	14,76 $\pm$ 0,39	1578 $\pm$ 19	0,7 $\pm$ 0,0	25,1 $\pm$ 2,0	72,5 $\pm$ 2	1,6 $\pm$ 0,0	2,92 $\pm$ 0,31
<i>f5hcp</i> 15-15	53,95 $\pm$ 2,85	15,01 $\pm$ 0,51	1520 $\pm$ 31	0,7 $\pm$ 0,0	35,1 $\pm$ 3,3	62,9 $\pm$ 3,2	1,2 $\pm$ 0,1	1,82 $\pm$ 0,26

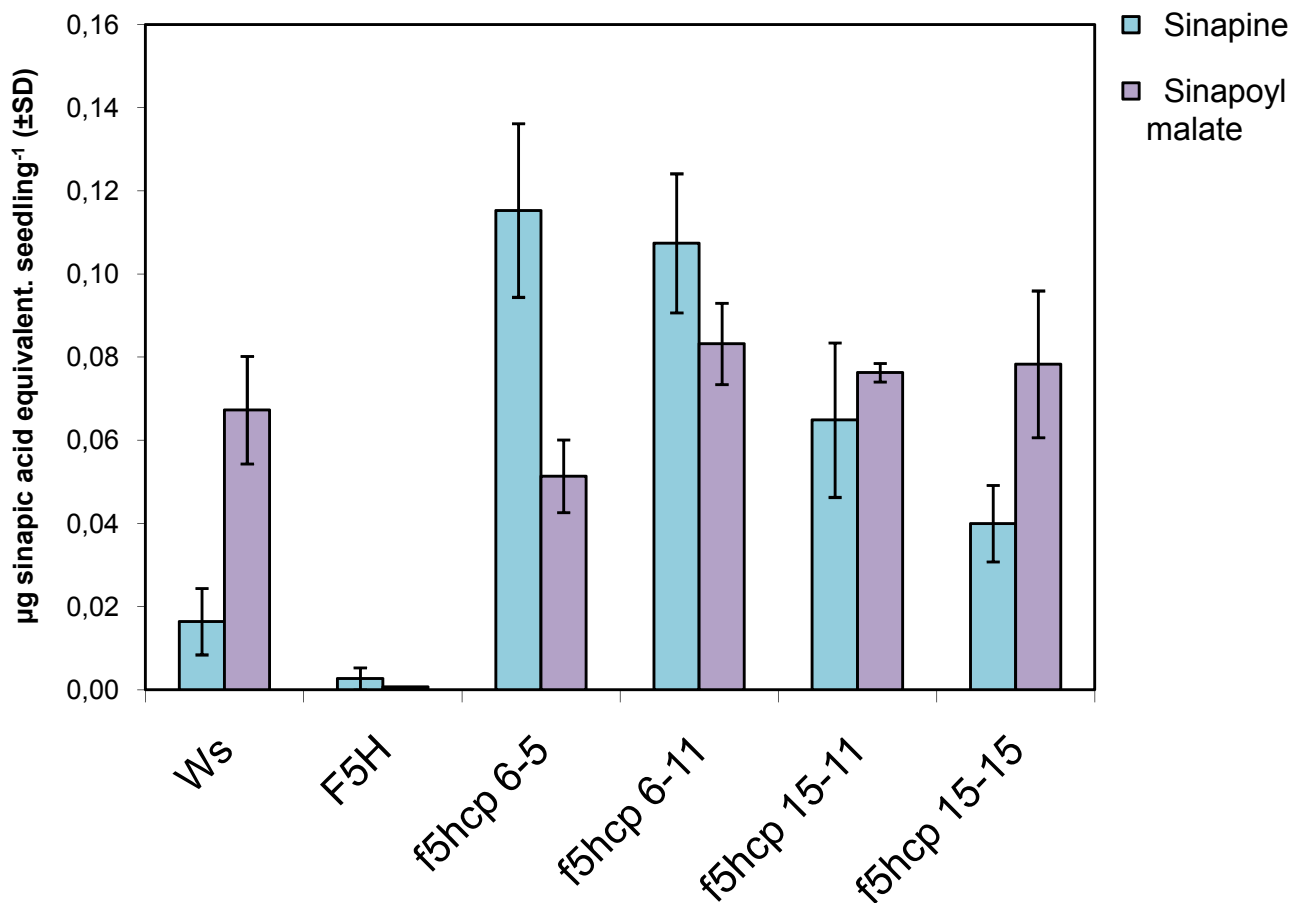
Ces résultats suggèrent que *BdF5H2* a bien une activité équivalente à celle d'*AtF5H1* dans la voie de biosynthèse des monolignols et les mutants *Atf5h* sont bien complémentés.

## 5.2. Composés phénoliques solubles

La perte de fonction de la F5H chez *Arabidopsis thaliana* a également des conséquences sur des composés pariétaux autres que les lignines : la diminution importante de la quantité de 5-hydroxyconiféraldéhyde cause une diminution de la quantité de sinapaldéhyde dans les cellules, qui est le composé suivant dans voie de biosynthèse des monolignols, après action de la COMT. Or le sinapaldéhyde est un composé essentiel dans la voie de biosynthèse du sinapate ester et son absence se répercute sur les quantités de sinapoyl choline (ou sinapine), de sinapate, de sinapoyl glucose et de sinapoyl malate. En accord avec les résultats déjà publiés (Chapple, Vogt et al. 1992; Ruegger, Meyer et al. 1999), ces dérivés sinapiques sont pratiquement absents chez les plantes mutantes pour le gène *F5H*. La sinapine et le sinapoyl malate ont été dosés sur des plantules de 10 jours cultivées in vitro sur boîtes de Pétri. La quantité de sinapoyl malate est restaurée dans les lignées complémentées, à des niveaux équivalents à celui du sauvage. Les quantités de sinapine dans les lignées complémentées sont plus importantes que chez le sauvage, suggérant l'existence d'au moins une étape limitante dans la voie du sinapoyl malate, située après formation de la sinapine et montrant la complémentation efficace de la lignée mutante dans la *F5H1* d'*Arabidopsis*.

## 6. Conclusion et perspectives

La convergence des données phylogénétiques, d'expression et fonctionnelles suggère le rôle de *BdF5H2* dans la voie de biosynthèse des monolignols chez *Brachypodium distachyon*.



**Figure 7.** Teneurs en sinapine et en sinapoyl malate de plantules T3 de 10 jours cultivées sur boîte de Pétri et correspondant au témoins (WS), au mutant *f5h* (mutant EBL87) et à 4 lignées d'*Arabidopsis thaliana* issues de la complémentation du mutant *f5h* par BdF5H2. Les résultats correspondent aux moyennes et écarts-types entre 3 répétitions et sont exprimés en équivalent d'acide sinapique (en µg) par plantule (récoltée fraîche).

Le TILLING n'a pas été commencé sur ce gène, mais il sera intéressant de caractériser des mutants pour ce gène dans le cadre de la recherche d'amélioration de rendement de saccharification. En effet, la perte de fonction du gène *F5H* impliqué dans la voie de biosynthèse des monolignols provoque une disparition presque complète des unités S des lignines, disparition associée à la diminution de la fréquence des liaisons  $\beta$ -O-4 dans ces polymères. A l'inverse, comme le suggèrent les résultats obtenus sur lignées complémentées, il est possible d'obtenir des lignines ayant un taux d'unités S proche de 90% en surexprimant ce gène, cette forte teneur en unités S étant associée à une fréquence plus élevée des liaisons  $\beta$ -O-4. Ainsi, en modulant l'activité F5H, il serait possible de moduler simultanément la fréquence des unités S et des liaisons  $\beta$ -O-4 dans les lignines. S'il est à présent bien établi que ces paramètres ont un impact positif sur l'aptitude des parois lignocellulosiques à être converties en pâte à papier par voie chimique, leurs effets sur la dégradabilité enzymatique des lignocelluloses dédiées à être converties en éthanol carburant est l'objet de résultats parfois contradictoires (Jung, Ni et al. 1999; Zhang, Culhaoglu et al. 2011).

## DISCUSSION GENERALE ET PERSPECTIVES





# Discussion générale et perspectives

## Intoduction

La nécessité de trouver des sources d'énergies renouvelables pouvant à long terme remplacer l'utilisation des carburants fossiles a soulevé l'intérêt pour les biocarburants de seconde génération, et en particulier le bioéthanol produit à partir des polysaccharides de la paroi lignocellulosique. La première étape de ce processus est la saccharification des parois, c'est-à-dire la dégradation enzymatique des sucres complexes que sont la cellulose et les hémicelluloses en sucres simples fermentescibles. Or cette étape est un facteur limitant des rendements de production de biocarburants. En effet, l'organisation des parois végétales constitue un obstacle à l'activité des enzymes, notamment en raison de la présence de lignines qui représentent un composant majeur de l'architecture pariétale et qui limitent l'accès aux polysaccharides. La biosynthèse des lignines chez les dicotylédones a été étudiée de façon approfondie chez les dicotylédones, et en particulier la plante modèle *Arabidopsis thaliana*, mais dans une moindre mesure chez les graminées. L'émergence de la nouvelle plante modèle *Brachypodium distachyon* est l'occasion de mettre en place des outils et des bases d'étude de la lignification chez les graminées, dans un organisme plus simple que les espèces d'intérêt agronomique sur lesquels le travail à grande échelle est difficile et qui présentent souvent un génome de très grande taille et polyploïde, générant beaucoup de redondances fonctionnelles.



## Validation de la démarche

L'objectif des travaux décrits consistait à identifier des gènes impliqués dans la voie de biosynthèse des monolignols pouvant être à la base de stratégies de modifications de l'organisation des parois végétales dans le but de faciliter leur saccharification.

La démarche employée pour remplir cet objectif a été d'exploiter la collection de mutants de *Brachypodium*, obtenue par mutagenèse chimique.

La stratégie de génétique classique employée sur les lignées *brown stem (bs)* s'est révélée efficace pour l'identification de lignées dont les rendements de saccharification sont meilleurs que ceux de la plante sauvage. Cependant l'identification du gène causant ce phénotype n'a été possible que dans le cas où, en plus du phénotype coloré visible, les lignines des mutants présentaient des marqueurs diagnostiques de sa déficience permettant son identification. Les mutations causant les phénotypes des autres membres de la famille *bs* sont encore à ce jour non identifiées et ne peuvent donc pas être appliquées aux espèces d'intérêt agronomique. Cependant, l'existence d'une grande variabilité génétique naturelle entre les accessions de *Brachypodium distachyon* pourrait permettre de cartographier ces mutations.

La stratégie de génétique inverse par TILLING a permis d'identifier un certain nombre de mutants qui ne présentent pas de phénotype visible. Or ce sont ces modifications d'architecture pariétale sans impact sur le développement des plantes qui présenteront un intérêt pour être transférées aux espèces chez lesquelles elles amélioreront les propriétés d'usage industriel pour lesquelles elles auront été sélectionnées. De plus, l'analyse d'un nombre plus important d'allèles mutants différents pourrait permettre d'identifier les sites essentiels à l'activité de protéines pour lesquelles des données de cristallisation ne sont pas disponibles. Ces allèles



peuvent également présenter une gamme de déficiences plus ou moins fortes et permettre l'étude de variations d'activité.

Cette stratégie trouve son origine dans la détermination de gènes candidats. Celle-ci a montré sa limite dans le cas de la *CCoAOMT* où les homologies de séquences ne sont pas nécessairement un gage de la fonction d'une protéine. Le développement de nouveaux outils tels que Bradinet permet de contourner ce type d'obstacle et suggère que la convergence de nombreux facteurs doit être prise en compte pour la définition d'un gène candidat. Des différences étant marquées entre les graminées et les dicotylédones, certains outils adaptés aux unes ne le seront pas aux secondes. L'utilisation de l'outil Diurnal a été une illustration de ces différences : les profils sont très similaires entre les différents gènes de biosynthèse des monolignols chez *Arabidopsis*, mais ceux des gènes étudiés chez *Brachypodium* ne présentent des profils similaires ni entre eux, ni avec ceux d'*Arabidopsis*. Cet outil contient également des gènes de riz qu'il aurait été intéressant de comparer à ceux de notre modèle, mais aucune donnée n'est disponible sur les orthologues recherchés.

Identification de gènes de la voie de biosynthèse et caractérisation des effets de leurs mutations chez *Brachypodium*

La similitude des gènes de la voie de biosynthèse des monolignols entre les graminées et les dicotylédones a été établie pour trois d'entre eux : *F5H*, *COMT* et *CAD*. Les orthologues de *Brachypodium* ont efficacement complété les mutants *Arabidopsis* correspondants, ce qui a pu être démontré à la fois sur les lignines et sur les composés phénoliques solubles.



Cependant, les parois des graminées présentent des spécificités pour lesquelles les études sur *Arabidopsis thaliana* ne permettent pas d'apporter de réponses. C'est d'ailleurs la raison pour laquelle le besoin d'un modèle d'étude plus adapté comme *Brachypodium* a été ressenti. Notamment, la présence d'acides hydroxycinnamiques, caractéristiques des parois des graminées, a une influence sur l'accessibilité des enzymes aux polysaccharides pariétaux et donc sur les rendements de saccharification. Si l'impact des esters féruliques pontant lignines et polysaccharides sur la dégradabilité enzymatique est étayé par de nombreux travaux, celui des esters *p*-coumariques principalement associés aux lignines est plus mystérieux. L'étude de ce composé et les mécanismes de son dépôt sur les polymères pariétaux est un axe de travail qui sera étudié de manière plus approfondie dans l'équipe paroi secondaire de l'IJPB, et les mutants *Bdcomt3* présentant un phénotype au niveau de leurs lignines et de l'acide *p*-coumarique, identifiés au cours de ce travail, constituent des modèles précieux pour ce faire.

#### Utilisation des mutants pour d'autres études

L'implication de deux gènes, *BdCAD1* et *BdCOMT3*, dans la voie de biosynthèse des monolignols a pu être montrée par des analyses fonctionnelles et la caractérisation de mutants. La présence d'autres mutations dans le génome des lignées analysés, due à l'action aléatoire de la mutagenèse chimique employée pour mettre au point la population de mutants, a été compensée par la caractérisation de plusieurs mutants dans le même gène et présentant des phénotypes similaires. Ces lignées mutantes peuvent servir à d'autres études aux enjeux importants, telle que l'étude de leur susceptibilité vis-à-vis des pathogènes. Pour qu'elles soient valides, il serait important de débarrasser les lignées des mutations qu'elles contiennent dans le reste de leur génome, afin de limiter les effets observés à ceux causés par la mutation dans le gène d'intérêt. Il est considéré qu'une série de cinq ou six rétrocroisements des lignées





mutantes avec des plantes sauvages permettrait d'obtenir un fond génétique satisfaisant où le seul effet observable serait dû à la mutation.

Exploitation des outils mis en place pendant ces travaux de thèse

La plateforme de TILLING sur *Brachypodium* est fonctionnelle et permettrait de trouver des mutants dans d'autres gènes de la voie de biosynthèse que ceux présentés ici, à commencer par *BdF5H2* et *BdCCoAOMT3* afin de terminer la validation du premier et vérifier la fonction du second. Par ailleurs, l'existence chez de nombreuses espèces de plusieurs isoformes des protéines de la voie de biosynthèse des monolignols suggère que *Brachypodium* posséderait lui aussi plusieurs protéines à l'activité similaire, possiblement redondante. Ceci expliquerait la diminution d'activité CAD relativement faible dans les mutants *Bdcad1*. La recherche par TILLING de mutants dans d'autres candidats, par exemple *BdCAD4* ou *BdCAD7* permettrait d'obtenir des doubles, voire des triples mutants si nécessaire, conduisant à réduire totalement l'activité CAD. Nous pourrions ainsi identifier tous les gènes *CAD* impliqués dans la biosynthèse des monolignols chez *Brachypodium*, plus facilement que chez d'autres espèces de graminées pour lesquels la recherche de mutants et leurs croisements seraient plus difficiles à mettre en œuvre. L'identification de familles de gènes chez *Brachypodium* permettrait alors de diriger les études vers des gènes candidats chez des graminées d'intérêt agronomique.



## Références bibliographiques

- Albenne, C., H. Canut, et al. (2009). "Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships." Molecular plant **2**(5): 977.
- Alves, S. C., B. Worland, et al. (2009). "A protocol for Agrobacterium-mediated transformation of *Brachypodium distachyon* community standard line Bd21." Nature protocols **4**(5): 638-649.
- Anterola, A. and N. Lewis (2002). "Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity." Phytochemistry **64**(3): 221-294.
- Anterola, A. M., H. van Rensburg, et al. (1999). "Multi-site modulation of flux during monolignol formation in loblolly pine (*Pinus taeda*)." Biochemical and biophysical research communications **264**(3): 652-657.
- Atanassova, R., N. Favet, et al. (1995). "Altered lignin composition in transgenic tobacco expressing O-methyltransferase sequences in sense and antisense orientation." The Plant Journal **8**(4): 465-477.
- Baker, D. and A. Sali (2001). "Protein structure prediction and structural genomics." Science **294**(5540): 93.
- Barakat, A., A. Bagniewska-Zadworna, et al. (2009). "The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression." BMC Plant Biology **9**(1): 26.
- Barceló, A. R. (1997). "Lignification in plant cell walls." International Review of Cytology **476**: 87-132.
- Barrière, Y., C. Riboulet, et al. (2007). "Genetics and genomics of lignification in grass cell walls based on maize as model species." Genes Genomes Genomics **4**: 133-156.
- Baskin, T. I. (2001). "On the alignment of cellulose microfibrils by cortical microtubules: a review and a model." Protoplasma **245**(1): 150-171.
- Baucher, M., M. Bernard-vailhé, et al. (1999). "Down-regulation of cinnamyl alcohol dehydrogenase in transgenic alfalfa (*Medicago sativa* L.) and the effect on lignin composition and digestibility." Plant Molecular Biology **39**(3): 437-447.
- Baucher, M., B. Chabbert, et al. (1996). "Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar." Plant physiology **442**(4): 1479.
- Bechtold, N. and G. Pelletier (1998). "In Planta Agrobacterium-Mediated Transformation of Adult *Arabidopsis thaliana* Plants by Vacuum Infiltration." Arabidopsis protocols **82**: 259.
- Bell, J., D. Lehmann, et al. (2004). "Transgenic down-regulation of caffeic acid O-methyltransferase (COMT) led to improved digestibility in tall fescue (*Festuca arundinacea*)." Functional plant biology **34**(3): 235-245.
- Berthet, S., N. Demont-Caulet, et al. (2011). "Disruption of LACCASE4 and 17 Results in Tissue-Specific Alterations to Lignification of *Arabidopsis thaliana* Stems." The Plant Cell Online **23**(3): 1124.
- Boerjan, W., J. Ralph, et al. (2003). "LIGNIN BIOSYNTHESIS." Annual Review of Plant Biology **54**(1): 519-546.
- Bose, S. K., R. C. Francis, et al. (2009). "Lignin content versus syringyl to guaiacyl ratio amongst poplars." Bioresource technology **400**(4): 1628-1633.

- Bout, S. and W. Vermerris (2003). "A candidate-gene approach to clone the sorghum Brown midrib gene encoding caffeic acid O-methyltransferase." Molecular Genetics and Genomics **269**(2): 205-214.
- Burton, R. A. and G. B. Fincher (2009). "(1, 3; 1, 4)- $\beta$ -D-Glucans in Cell Walls of the Poaceae, Lower Plants, and Fungi: A Tale of Two Linkages." Molecular plant **2**(5): 873.
- Carpita, N. C. (1996). "Structure and biogenesis of the cell walls of grasses." Annual Review of Plant Biology **47**(1): 445-476.
- Cassab, G. I. and J. E. Varner (1988). "Cell wall proteins." Annual Review of Plant Physiology and Plant Molecular Biology **39**(1): 321-353.
- Catalán, P., E. A. Kellogg, et al. (1997). "Phylogeny of Poaceae Subfamily Pooideae Based on ChloroplastndhF Gene Sequences\* 1." Molecular Phylogenetics and Evolution **8**(2): 150-166.
- Catalán, P. and R. G. Olmstead (2000). "Phylogenetic reconstruction of the genus *Brachypodium* P. Beauv. (Poaceae) from combined sequences of chloroplastndhF gene and nuclear ITS." Plant Systematics and Evolution **220**(1): 1-19.
- Chabannes, M., A. Barakate, et al. (2001). "Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants." Plant J **28**(3): 257-270.
- Chapple, C., M. Ladisch, et al. (2007). "Loosening lignin's grip on biofuel production." Nature Biotechnology **25**(7): 746-748.
- Chapple, C. C. S., T. Vogt, et al. (1992). "An Arabidopsis mutant defective in the general phenylpropanoid pathway." The Plant Cell Online **4**(11): 1413-1424.
- Chen, F. and R. Dixon (2007). "Lignin modification improves fermentable sugar yields for biofuel production."
- Chen, F., M. S. Srinivasa Reddy, et al. (2006). "Multi-site genetic modulation of monolignol biosynthesis suggests new routes for formation of syringyl lignin and wall-bound ferulic acid in alfalfa (*Medicago sativa* L.)." The Plant Journal **48**(1): 113-124.
- Chen, F., S. Yasuda, et al. (1999). "Evidence for a novel biosynthetic pathway that regulates the ratio of syringyl to guaiacyl residues in lignin in the differentiating xylem of *Magnolia kobus* DC." Planta **207**(4): 597-603.
- Cherney, D., D. Mertens, et al. (1990). "Intake and digestibility by wethers as influenced by forage morphology at three levels of forage offering." Journal of animal science **68**(12): 4387.
- Cherney, J., D. Cherney, et al. (1991). "Potential of brown-midrib, low-lignin mutants for improving forage quality." Advances in Agronomy **46**: 157-198.
- Civardi, L., J. Rigau, et al. (1999). "Nucleotide sequence of two cDNAs coding for caffeoyl-coenzyme A O-methyltransferase (CCoAOMT) and study of their expression in *Zea mays*." Plant Physiol **420**(4): 1.
- Cochrane, F. C., L. B. Davin, et al. (2004). "The Arabidopsis phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms." Phytochemistry **65**(11): 1557-1564.
- Cosgrove, D. J. (1999). "Enzymes and other agents that enhance cell wall extensibility." Annual Review of Plant Biology **50**(1): 391-417.
- Cosgrove, D. J. (2005). "Growth of the plant cell wall." Nature Reviews Molecular Cell Biology **6**(11): 850-861.
- Crowell, E. F., M. Gonneau, et al. (2010). "Regulation of anisotropic cell expansion in higher plants." Comptes Rendus Biologies **333**(4): 320-324.

- Cymbaluk, N. F., A. Gordon, et al. (1973). "The effect of the chemical composition of maize plant lignin on the digestibility of maize stalk in the rumen of cattle." British Journal of Nutrition **29**(01): 1-12.
- D'Ovidio, R., B. Mattei, et al. (2004). "Polygalacturonases, polygalacturonase-inhibiting proteins and pectic oligomers in plant-pathogen interactions." Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics **4696**(2): 237-244.
- Damiani, I., K. Morreel, et al. (2005). "Metabolite profiling reveals a role for atypical cinnamyl alcohol dehydrogenase CAD1 in the synthesis of coniferyl alcohol in tobacco xylem." Plant molecular biology **59**(5): 753-769.
- Dean, J. F. D. and K. E. L. Eriksson (1994). "Laccase and the deposition of lignin in vascular plants." Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood **48**(s1): 21-33.
- Dèjardin, A., F. Laurans, et al. (2010). "Wood formation in Angiosperms." Comptes Rendus Biologies **333**(4): 325-334.
- Demont-Caulet, N., C. Lapierre, et al. (2010). "Arabidopsis peroxidase-catalyzed copolymerization of coniferyl and sinapyl alcohols: Kinetics of an endwise process." Phytochemistry **74**(14-15): 1673-1683.
- Dence, C. (1992). "Lignin determination." Methods in Lignin Chemistry. Springer-Verlag, Berlin: 33ñ61.
- Dharmawardhana, D. P., B. E. Ellis, et al. (1995). "A [beta]-Glucosidase from Lodgepole Pine Xylem Specific for the Lignin Precursor Coniferin." Plant physiology **407**(2): 331-339.
- Dien, B. S., G. Sarath, et al. (2009). "Improved sugar conversion and ethanol yield for forage sorghum (*Sorghum bicolor* L. Moench) lines with reduced lignin contents." Bioenergy Research **2**(3): 153-164.
- Do, C. T., B. Pollet, et al. (2007). "Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis." Planta **226**(5): 1117-1129.
- Donaldson, L. A. (2001). "Lignification and lignin topochemistry--an ultrastructural view." Phytochemistry **57**(6): 859-873.
- Draper, J. (2001). "Brachypodium distachyon. A new model system for functional genomics in grasses." Plant Physiol. **427**: 1539-1555.
- Ehlting, J., N. Mattheus, et al. (2005). "Global transcript profiling of primary stems from Arabidopsis thaliana identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation." The Plant Journal **42**(5): 618-640.
- Escamilla Treviño, L. L., H. Shen, et al. (2010). "Switchgrass (*Panicum virgatum*) possesses a divergent family of cinnamoyl CoA reductases with distinct biochemical properties." New Phytologist **485**(1): 143-155.
- Escamilla-Treviño, L. L., W. Chen, et al. (2006). "Arabidopsis thaliana [beta]-Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides." Phytochemistry **67**(15): 1651-1660.
- Fergus, B. and D. Goring (1970). "The distribution of lignin in birch wood as determined by ultraviolet microscopy." Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood **24**(4): 118-124.
- Fincher, G. B. (2009). "Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses." Plant physiology **449**(1): 27.
- Ford, C., R. Elliott, et al. (1987). "The effect of chlorite delignification on digestibility of some grass forages and on intake and rumen microbial activity in sheep fed barley straw." The Journal of Agricultural Science **408**(01): 129-136.

- Fornalé, S., F. M. Sonbol, et al. (2006). "Down-regulation of the maize and Arabidopsis thaliana caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors." Plant Molecular Biology **62**(6): 809-823.
- Franke, R., C. M. McMichael, et al. (2000). "Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase." The Plant Journal **22**(3): 223-234.
- Fu, C., J. R. Mielenz, et al. (2011). "Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass." Proceedings of the National Academy of Sciences **408**(9): 3803.
- Fu, C., X. Xiao, et al. (2011). "Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass." Bioenergy Research: 1-12.
- Fukushima, K. and N. Terashima (1991). "Heterogeneity in formation of lignin." Wood science and technology **25**(5): 371-381.
- Gang, D. R., M. A. Costa, et al. (1999). "Regiochemical control of monolignol radical coupling: a new paradigm for lignin and lignan biosynthesis." Chemistry & biology **6**(3): 143-151.
- Gardiner, J. C., N. G. Taylor, et al. (2003). "Control of cellulose synthase complex localization in developing xylem." The Plant Cell Online **45**(8): 1740.
- Goffner, D., M. M. Campbell, et al. (1994). "Purification and characterization of cinnamoyl-coenzyme A: NADP oxidoreductase in Eucalyptus gunnii." Plant physiology **406**(2): 625.
- Goffner, D., I. Joffroy, et al. (1992). "Purification and characterization of isoforms of cinnamyl alcohol dehydrogenase from Eucalyptus xylem." Planta **488**(1): 48-53.
- Gomez, L. D., C. Whitehead, et al. (2010). "Automated saccharification assay for determination of digestibility in plant materials." Biotechnology for Biofuels **3**(1): 1-12.
- Goujon, T., V. Ferret, et al. (2003). "Down-regulation of the AtCCR1 gene in Arabidopsis thaliana: effects on phenotype, lignins and cell wall degradability." Planta **247**(2): 218-228.
- Goujon, T., R. Sibout, et al. (2003). "A new Arabidopsis thaliana mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters." Plant Molecular Biology **54**(6): 973-989.
- Grabber, J. H. (2005). "How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies." Crop Sci **45**(3): 820-831.
- Grabber, J. H. and G. A. Jung (1991). "In-vitro disappearance of carbohydrates, phenolic acids, and lignin from parenchyma and sclerenchyma cell walls isolated from cocksfoot." Journal of the Science of Food and Agriculture **57**(3): 315-323.
- Grabber, J. H., S. Quideau, et al. (1996). "p-coumaroylated syringyl units in maize lignin: Implications for [beta]-ether cleavage by thioacidolysis." Phytochemistry **43**(6): 1189-1194.
- Grabber, J. H., J. Ralph, et al. (2004). "Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions\* 1." Comptes Rendus Biologies **327**(5): 455-465.
- Grima-Pettenati, J., C. Feuillet, et al. (1993). "Molecular cloning and expression of a Eucalyptus gunnii cDNA clone encoding cinnamyl alcohol dehydrogenase." Plant Molecular Biology **24**(6): 1085-1095.
- Gritsch, C. S. and R. J. Murphy (2005). "Ultrastructure of fibre and parenchyma cell walls during early stages of culm development in Dendrocalamus asper." Annals of Botany **95**(4): 619.

- Guillon, F., B. Bouchet, et al. (2011). "Brachypodium distachyon grain: characterization of endosperm cell walls." Journal of experimental botany **62**(3): 1001.
- Guo, D., F. Chen, et al. (2001). "Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin." The Plant Cell Online **43**(1): 73-88.
- Halpin, C., K. Holt, et al. (1998). "Brown-midrib maize (bm1)-a mutation affecting the cinnamyl alcohol dehydrogenase gene." Plant Journal **44**(5): 545.
- Halpin, C., M. E. Knight, et al. (1992). "Purification and characterization of cinnamyl alcohol dehydrogenase from tobacco stems." Plant physiology **98**(1): 12.
- Harholt, J., A. Suttangkakul, et al. (2010). "Biosynthesis of pectin." Plant physiology **453**(2): 384.
- Harris, D., V. Bulone, et al. (2010). "Tools for cellulose analysis in plant cell walls." Plant physiology **453**(2): 420.
- Harris, P. J. and J. A. K. Trethewey (2010). "The distribution of ester-linked ferulic acid in the cell walls of angiosperms." Phytochemistry Reviews **9**(1): 19-33.
- Hartley, R. D., W. H. Morrison III, et al. (1992). "Phenolic constituents of cell wall types of normal and brown midrib mutants of pearl millet (*Pennisetum glaucum* (L) R Br) in relation to wall biodegradability." Journal of the Science of Food and Agriculture **59**(2): 211-216.
- Hasterok, R., J. Draper, et al. (2004). "Laying the cytotoxic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv." Chromosome Research **42**(4): 397-403.
- Hasterok, R., A. Marasek, et al. (2006). "Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in situ hybridization." Genetics **473**(1): 349.
- Hatfield, R., J. Ralph, et al. (1999). "Cell wall structural foundations: Molecular basis for improving forage digestibilities." Crop Sci **39**(1): 27-37.
- He, L. and N. Terashima (1990). "Formation and structure of lignin in monocotyledons. III. Heterogeneity of sugarcane (*Saccharum officinarum* L.) lignin with respect to the composition of structural units in different morphological regions." Journal of wood chemistry and technology **40**(4): 435-459.
- He, L. and N. Terashima (1991). "Formation and Structure of Lignin in Monocotyledons IV. Deposition Process and Structural Diversity of the Lignin in the CellWall of Sugarcane and Rice Plant Studied by Ultraviolet Microscopic Spectroscopy." Holzforchung-International Journal of the Biology, Chemistry, Physics and Technology of Wood **45**(3): 191-198.
- Higuchi, T. (1990). "Lignin biochemistry: biosynthesis and biodegradation." Wood science and technology **24**(1): 23-63.
- Higuchi, T., T. Ito, et al. (1994). "Red-brown color of lignified tissues of transgenic plants with antisense CAD gene: Wine-red lignin from coniferyl aldehyde." Journal of biotechnology **37**(2): 151-158.
- Himmelbach, A., U. Zierold, et al. (2007). "A set of modular binary vectors for transformation of cereals." Plant physiology **445**(4): 1192.
- Hoffmann, L., S. Besseau, et al. (2004). "Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis." The Plant Cell Online **46**(6): 1446.
- Holst Christensen, J., S. Overney, et al. (2001). "The syringaldazine-oxidizing peroxidase PXP 3-4 from poplar xylem: cDNA isolation, characterization and expression." Plant Molecular Biology **47**(5): 581-593.

- Hong, S.-Y., P. Seo, et al. (2008). "Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR." *BMC Plant Biology* **8**(1): 112.
- Humphreys, J. M., M. R. Hemm, et al. (1999). "New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase." *Proceedings of the National Academy of Sciences* **96**(18): 10045.
- Huo, G., G. Jiang, et al. (2007). "Phylogenetic reconstruction of the family acrypteridae (orthoptera: acridoidea) based on mitochondrial cytochrome B gene." *Journal of Genetics and Genomics* **34**(4): 294-306.
- Huo, N., G. R. Lazo, et al. (2008). "The nuclear genome of *Brachypodium distachyon*: analysis of BAC end sequences." *Functional & Integrative Genomics* **8**(2): 135-147.
- Huo, N., J. P. Vogel, et al. (2009). "Structural characterization of *Brachypodium* genome and its syntenic relationship with rice and wheat." *Plant Molecular Biology* **70**(1): 47-61.
- Iiyama, K., T. B. T. Lam, et al. (1990). "Phenolic acid bridges between polysaccharides and lignin in wheat internodes." *Phytochemistry* **29**(3): 733-737.
- Inoue, K., V. J. H. Sewalt, et al. (1998). "Developmental expression and substrate specificities of alfalfa caffeic acid 3-O-methyltransferase and caffeoyl coenzyme A 3-O-methyltransferase in relation to lignification." *Plant physiology* **447**(3): 761.
- Inoue, K., V. J. H. Sewalt, et al. (1998). "Developmental Expression and Substrate Specificities of Alfalfa Caffeic Acid 3-O-Methyltransferase and Caffeoyl Coenzyme A 3-O-Methyltransferase in Relation to Lignification." *Plant physiology* **447**(3): 761-770.
- Ishii, T. (1997). "Structure and functions of feruloylated polysaccharides." *Plant Science* **427**(2): 111-127.
- Jacquet, G., B. Pollet, et al. (1995). "New ether-linked ferulic acid-coniferyl alcohol dimers identified in grass straws." *Journal of agricultural and food chemistry* **43**(10): 2746-2751.
- Jamet, E., C. Albenne, et al. (2008). "Recent advances in plant cell wall proteomics." *Proteomics* **8**(4): 893-908.
- Jeoh, T., C. I. Ishizawa, et al. (2007). "Cellulase digestibility of pretreated biomass is limited by cellulose accessibility." *Biotechnology and bioengineering* **98**(1): 112-122.
- Joshi, C. P., S. Bhandari, et al. (2004). "Genomics of cellulose biosynthesis in poplars." *New Phytologist* **464**(1): 53-61.
- Jouanin, L., T. Goujon, et al. (2000). "Lignification in transgenic poplars with extremely reduced caffeic acid O-methyltransferase activity." *Plant physiology* **423**(4): 1363.
- Jung, H. J. G., W. Ni, et al. (1999). "Impact of lignin composition on cell wall degradability in an *Arabidopsis* mutant." *Journal of the Science of Food and Agriculture* **79**(6): 922-928.
- Kaneda, M., K. H. Rensing, et al. (2008). "Tracking monolignols during wood development in lodgepole pine." *Plant physiology* **447**(4): 1750.
- Kim, H., J. Ralph, et al. (2002). "Identification of the structure and origin of thioacidolysis marker compounds for cinnamyl alcohol dehydrogenase deficiency in angiosperms." *Journal of Biological Chemistry* **277**(49): 47412.
- Kuc, J., O. Nelson, et al. (1968). "Degradation of abnormal lignins in the brown-midrib mutants and double mutants of maize." *Phytochemistry* **7**(8): 1435-1436.
- Kuc, J. and O. E. Nelson (1964). "The abnormal lignins produced by the brown-midrib mutants of maize\* 1: I. The brown-midrib-1 mutant." *Archives of biochemistry and biophysics* **405**(1): 103-113.
- Lanot, A., D. Hodge, et al. (2006). "The glucosyltransferase UGT72E2 is responsible for monolignol 4 O glucoside production in *Arabidopsis thaliana*." *The Plant Journal* **48**(2): 286-295.



- Lapierre, C. (1993). "Application of new methods for the investigation of lignin structure." Forage cell wall structure and digestibility **4993**: 133-163.
- Lapierre, C., B. Monties, et al. (1985). "Thioacidolysis of lignin: comparison with acidolysis." Journal of wood chemistry and technology **5**(2): 277-292.
- Lapierre, C., G. Pilate, et al. (2004). "Signatures of cinnamyl alcohol dehydrogenase deficiency in poplar lignins." Phytochemistry **65**(3): 313-321.
- Lapierre, C., B. Pollet, et al. (1995). "New insights into the molecular architecture of hardwood lignins by chemical degradative methods." Research on chemical intermediates **24**(3): 397-412.
- Larsen, K. (2004). "Cloning and characterization of a ryegrass (*Lolium perenne*) gene encoding cinnamoyl-CoA reductase (CCR)." Plant Science **466**(3): 569-581.
- Laudencia-Chingcuanco, D. L. and W. H. Vensel (2008). "Globulins are the main seed storage proteins in *Brachypodium distachyon*." TAG Theoretical and Applied Genetics **447**(4): 555-563.
- Lauvergeat, V., C. Lacomme, et al. (2001). "Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria." Phytochemistry **57**(7): 1187-1195.
- Li, X., J. K. Weng, et al. (2008). "Improvement of biomass through lignin modification." Plant J **54**(4): 569-581.
- Li, X., Y. Yang, et al. (2009). "FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice." Plant Molecular Biology **69**(6): 685-697.
- Lim, E., Y. Li, et al. (2001). "Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in *Arabidopsis*." Journal of Biological Chemistry **276**(6): 4344-4349.
- Lim, E. K., R. G. Jackson, et al. (2005). "Identification and characterisation of *Arabidopsis* glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde." FEBS letters **579**(13): 2802-2806.
- Lin, F., G. Yamano, et al. (2006). "Cloning and functional analysis of caffeic acid 3-O-methyltransferase from rice (*Oryza sativa*)." Journal of Pesticide Science **34**(1): 47-53.
- Louie, G. V., M. E. Bowman, et al. (2010). "Structure-Function Analyses of a Caffeic Acid O-Methyltransferase from Perennial Ryegrass Reveal the Molecular Basis for Substrate Preference." The Plant Cell Online **22**(12): 4114.
- Lowry, B., D. Lee, et al. (1980). "The origin of land plants: a new look at an old problem." Taxon: 183-197.
- Lu, F. and J. Ralph (2002). "Preliminary evidence for sinapyl acetate as a lignin monomer in kenaf." Chemical Communications(1): 90-91.
- Lynd, L. R., J. H. Cushman, et al. (1991). "Fuel ethanol from cellulosic biomass." Science **254**(4999): 1318.
- Ma, J. and N. Yamaji (2008). "Functions and transport of silicon in plants." Cellular and Molecular Life Sciences **65**(19): 3049-3057.
- Ma, Q. H. (2007). "Characterization of a cinnamoyl-CoA reductase that is associated with stem development in wheat." Journal of experimental botany **58**(8): 2011.
- Ma, Q. H. and B. Tian (2005). "Biochemical characterization of a cinnamoyl-CoA reductase from wheat." Biological chemistry **386**(6): 553-560.
- MacKay, J. J., D. M. O'Malley, et al. (1997). "Inheritance, gene expression, and lignin characterization in a mutant pine deficient in cinnamyl alcohol dehydrogenase." Proceedings of the National Academy of Sciences **94**(15): 8255.

- Marita, J. M., J. Ralph, et al. (1999). "NMR characterization of lignins in Arabidopsis altered in the activity of ferulate 5-hydroxylase." Proceedings of the National Academy of Sciences **96**(22): 12328.
- Martone, P. T., J. M. Estevez, et al. (2009). "Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture." Current Biology **49**(2): 169-175.
- McCallum, C. M., L. Comai, et al. (2000). "Targeting induced local lesions in genomes (TILLING) for plant functional genomics." Plant physiology **423**(2): 439.
- McInnes, R., A. Lidgett, et al. (2002). "Isolation and characterization of a cinnamoyl-CoA reductase gene from perennial ryegrass (*Lolium perenne*)." Journal of plant physiology **459**(4): 415-422.
- Meyer, K., J. Cusumano, et al. (1997). Regulation and manipulation of lignin monomer composition by expression and over-expression of ferulate 5-hydroxylase, a cytochrome P450-dependent monooxygenase required for syringyl lignin biosynthesis.
- Meyermans, H., K. Morreel, et al. (2000). "Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis." Journal of Biological Chemistry **275**(47): 36899.
- Mir Derikvand, M., J. Sierra, et al. (2008). "Redirection of the phenylpropanoid pathway to feruloyl malate in Arabidopsis mutants deficient for cinnamoyl-CoA reductase 1." Planta **227**(5): 943-956.
- Mockler, T., T. Michael, et al. (2007). "The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis." Clocks and rhythms **72**: 353.
- Mohnen, D. (2008). "Pectin structure and biosynthesis." Current opinion in plant biology **44**(3): 266-277.
- Morrison III, W. and D. Archibald (1998). "Analysis of graded flax fiber and yarn by pyrolysis mass spectrometry and pyrolysis gas chromatography mass spectrometry." Journal of agricultural and food chemistry **46**(5): 1870-1876.
- Musel, G., T. Schindler, et al. (1997). "Distribution of lignin primary and secondary cell wall of maize coleoptiles analyzed by chemical and immunology probes." Planta **204**: 146-159.
- Nagaraj, S. H., R. B. Gasser, et al. (2007). "A hitchhiker's guide to expressed sequence tag (EST) analysis." Briefings in Bioinformatics **8**(1): 6-21.
- Ng, P. C. and S. Henikoff (2006). "Predicting the effects of amino acid substitutions on protein function." Annu. Rev. Genomics Hum. Genet. **7**: 61-80.
- O'Malley, D. M., S. Porter, et al. (1992). "Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in loblolly pine (*Pinus taeda* L.)." Plant physiology **98**(4): 1364.
- Opanowicz, M., P. Vain, et al. (2008). "Brachypodium distachyon: making hay with a wild grass." Trends in Plant Science **43**(4): 172-177.
- Osakabe, K., C. Tsao, et al. (1999). Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms, National Acad Sciences. **96**: 8955-8960.
- Parvathi, K., F. Chen, et al. (2001). "Substrate preferences of O methyltransferases in alfalfa suggest new pathways for 3 O methylation of monolignols." The Plant Journal **25**(2): 193-202.
- Pauly, M., P. Albersheim, et al. (1999). "Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants." The Plant Journal **20**(6): 629-639.
- Pichon, M., I. Courbou, et al. (1998). "Cloning and characterization of two maize cDNAs encoding Cinnamoyl-CoA Reductase (CCR) and differential expression of the corresponding genes." Plant Molecular Biology **38**(4): 671-676.

- Pilling, E. and H. Hofte (2003). "Feedback from the wall." Current opinion in plant biology **6**(6): 611-616.
- Pinçon, G., S. Maury, et al. (2001). "Repression of O-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth." Phytochemistry **57**(7): 1167-1176.
- Piquemal, J., S. Chamayou, et al. (2002). "Down-regulation of caffeic acid O-methyltransferase in maize revisited using a transgenic approach." Plant Physiology **430**(4): 1675-1685.
- Piquemal, J., C. Lapierre, et al. (1998). "Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants." The Plant Journal **43**(1): 71-83.
- Quraishi, U. M., F. Murat, et al. (2011). "Combined meta-genomics analyses unravel candidate genes for the grain dietary fiber content in bread wheat (*Triticum aestivum* L.)." Functional & Integrative Genomics: 1-13.
- Raes, J., A. Rohde, et al. (2003). Genome-Wide Characterization of the Lignification Toolbox in *Arabidopsis* 1 [w], *Am Soc Plant Biol.* 133: 1051-1071.
- Ralph, J., J. H. Grabber, et al. (1995). "Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins\* 1." Carbohydrate research **275**(1): 167-178.
- Ralph, J., R. D. Hatfield, et al. (1994). "Pathway of *p*-coumaric acid incorporation into maize lignin as revealed by NMR." Journal of the American Chemical Society **446**(21): 9448-9456.
- Ralph, J., C. Lapierre, et al. (2001). "Elucidation of new structures in lignins of CAD-and COMT-deficient plants by NMR." Phytochemistry **57**(6): 993-1003.
- Ralph, J., K. Lundquist, et al. (2004). "Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids." Phytochemistry Reviews **3**(1): 29-60.
- Ranocha, P., M. Chabannes, et al. (2002). "Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar." Plant physiology **429**(1): 145-155.
- Richardson, A., N. Deighton, et al. (2000). "Oxidases participate in the formation of wood." SCRI Annual Report. pp129-131.
- Rosler, J., F. Krekel, et al. (1997). "Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity." Plant physiology **443**(1): 175.
- Rossmann, M. G., D. Moras, et al. (1974). "Chemical and biological evolution of a nucleotide-binding protein." Nature **250**: 194-199.
- Ruegger, M., K. Meyer, et al. (1999). "Regulation of ferulate-5-hydroxylase expression in *Arabidopsis* in the context of sinapate ester biosynthesis." Plant physiology **449**(1): 101-110.
- Ryden, P., K. Sugimoto-Shirasu, et al. (2003). "Tensile properties of *Arabidopsis* cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II-borate complexes." Plant physiology **432**(2): 1033.
- Sánchez-Rodríguez, C., I. Rubio-Somoza, et al. (2010). "Phytohormones and the cell wall in *Arabidopsis* during seedling growth." Trends in plant science **45**(5): 291-301.
- Saathoff, A. J., C. M. Tobias, et al. (2011). "Switchgrass contains two cinnamyl alcohol dehydrogenases involved in lignin formation." Bioenergy Research: 1-14.
- Saballos, A., G. Ejeta, et al. (2009). "A Genomewide Analysis of the Cinnamyl Alcohol Dehydrogenase Family in *Sorghum* [*Sorghum bicolor* (L.) Moench] Identifies SbCAD2 as the Brown midrib6 Gene." Genetics **484**(2): 783-795.
- Saballos, A., W. Vermerris, et al. (2008). "Allelic association, chemical characterization and saccharification properties of brown midrib mutants of sorghum (*Sorghum bicolor* (L.) Moench)." Bioenergy Research **4**(3): 193-204.

- Samuels, A., K. Rensing, et al. (2002). "Cellular machinery of wood production: differentiation of secondary xylem in *Pinus contorta* var. *latifolia*." Planta **246**(1): 72-82.
- Sato, Y., T. Watanabe, et al. (1997). "Changes in the activity and mRNA of cinnamyl alcohol dehydrogenase during tracheary element differentiation in *Zinnia*." Plant physiology **443**(2): 425.
- Sato, Y., B. Wuli, et al. (2001). "Molecular Cloning and Expression of Eight Laccase cDNAs in Loblolly Pine (*Pinus taeda*)". Journal of Plant Research **444**(2): 147-155.
- Sattler, S. E., A. J. Saathoff, et al. (2009). "A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the sorghum brown midrib6 phenotype." Plant physiology **450**(2): 584.
- Scheller, H. V. and P. Ulvskov (2010). "Hemicelluloses." Annual Review of Plant Biology **64**: 263-289.
- Schoch, G. A., M. Morant, et al. (2006). "The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation." Environmental Chemistry Letters **4**(3): 127-136.
- Selman-Housein, G. and M. I. Lùpez (1999). "Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification." Plant Science **443**(2): 163-171.
- Shane, M., M. McCully, et al. (2000). "The vascular system of maize stems revisited: implications for water transport and xylem safety." Annals of Botany **86**(2): 245.
- Showalter, A. (2001). "Introduction: plant cell wall proteins." Cellular and Molecular Life Sciences **58**(10): 1361-1362.
- Sibout, R., A. Eudes, et al. (2005). "CINNAMYL ALCOHOL DEHYDROGENASE-C and-D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*." The Plant Cell Online **47**(7): 2059-2076.
- Sibout, R., A. Eudes, et al. (2003). "Expression Pattern of Two Paralogs Encoding Cinnamyl Alcohol Dehydrogenases in *Arabidopsis*. Isolation and Characterization of the Corresponding Mutants 1." Plant Physiology **432**(2): 848-860.
- Simmons, B. A., D. Loqu , et al. (2010). "Advances in modifying lignin for enhanced biofuel production." Current opinion in plant biology **43**(3): 312-319.
- Smith, B. G. and P. J. Harris (1999). "The polysaccharide composition of Poales cell walls:: Poaceae cell walls are not unique." Biochemical Systematics and Ecology **27**(1): 33-53.
- Somers, D. A., J. P. Nourse, et al. (1995). "A gene encoding a cinnamyl alcohol dehydrogenase homolog in *Arabidopsis thaliana*." Plant physiology **408**(3): 1309.
- Somerville, C., S. Bauer, et al. (2004). "Toward a systems approach to understanding plant cell walls." Science **306**(5705): 2206.
- Somerville, C., H. Youngs, et al. (2010). "Feedstocks for lignocellulosic biofuels." Science **329**(5993): 790.
- Stewart, J. J., T. Akiyama, et al. (2009). "The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar1." Plant physiology **450**(2): 621.
- Tamasloukht, B., M. S. J. Wong Quai Lam, et al. (2011). "Characterization of a cinnamoyl-CoA reductase 1 (CCR1) mutant in maize: effects on lignification, fibre development, and global gene expression." Journal of experimental botany **62**(11): 3837.
- Tanaka, T. (2008). "The rice annotation project database (RAP-DB): 2008 update." Nucleic Acids Res. **36**: D1028-D1033.
- Taylor, N. and S. Turner (2007). "Cellulose synthesis in the *Arabidopsis* secondary cell wall." Cellulose: Molecular and Structural Biology: 49-61.
- Tobias, C. and E. Chow (2005). "Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification." Planta **220**(5): 678-688.

- Tobias, C. M. and E. K. Chow (2005). "Structure of the cinnamyl-alcohol dehydrogenase gene family in rice and promoter activity of a member associated with lignification." Planta **220**(5): 678-688.
- Tu, Y., S. Rochfort, et al. (2010). "Functional analyses of caffeic acid O-methyltransferase and cinnamoyl-CoA-reductase genes from perennial ryegrass (*Lolium perenne*)." The Plant Cell Online **22**(10): 3357.
- Urbanowicz, B. R., C. Rayon, et al. (2004). "Topology of the Maize Mixed Linkage (1=> BORDER=" 0"> 3),(1=> BORDER=" 0"> 4)- -D-Glucan Synthase at the Golgi Membrane." Plant physiology **434**(2): 758.
- Vailhé, M. A. B., G. J. Provan, et al. (2000). "Effect of phenolic structures on the degradability of cell walls isolated from newly extended apical internode of tall fescue (*Festuca arundinacea* Schreb.)." Journal of agricultural and food chemistry **48**(3): 618-623.
- Vain, P., B. Worland, et al. (2008). "Agrobacterium-mediated transformation of the temperate grass *Brachypodium distachyon* (genotype Bd21) for T-DNA insertional mutagenesis." Plant Biotechnol J **6**: 236 - 245.
- Vermerris, W., D. M. Sherman, et al. (2010). "Phenotypic plasticity in cell walls of maize brown midrib mutants is limited by lignin composition." Journal of experimental botany **64**(9): 2479.
- Vignols, F., J. Rigau, et al. (1995). "The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase." The Plant Cell Online **7**(4): 407-416.
- Vogel, J. (2008). "Unique aspects of the grass cell wall." Curr. Opin. Plant Biol. **44**: 301-307.
- Vogel, J. and J. Bragg (2009). "*Brachypodium distachyon*, a new model for the Triticeae." Genetics and genomics of the Triticeae: 427-449.
- Vogel, J., D. Garvin, et al. (2006). "Agrobacterium-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*." Plant Cell, Tissue and Organ Culture **84**(2): 100179-100191.
- Vogel, J., Y. Gu, et al. (2006). "EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*." Theor Appl Genet **443**: 186 - 195.
- Vogel, J. and T. Hill (2008). "High-efficiency Agrobacterium-mediated transformation of *Brachypodium distachyon* inbred line Bd21-3." Plant cell reports **27**(3): 471-478.
- Vogel, J. P., D. F. Garvin, et al. (2010). "Genome sequencing and analysis of the model grass *Brachypodium distachyon*." Nature **463**(7282): 763-768.
- Wang, X., H. Tang, et al. (2011). "Seventy Million Years of Concerted Evolution of a Homoeologous Chromosome Pair, in Parallel, in Major Poaceae Lineages." The Plant Cell Online **23**(1): 27.
- Wei, J. H., Y. Z. Wang, et al. (2008). "Pulping performance of transgenic poplar with depressed caffeoyl-CoA O-methyltransferase." Chinese Science Bulletin **53**(22): 3553-3558.
- Welinder, K. G., A. F. Justesen, et al. (2002). "Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*." European Journal of Biochemistry **269**(24): 6063-6081.
- Weng, J. K., J. A. Banks, et al. (2008). "Parallels in lignin biosynthesis: a study in *Selaginella moellendorffii* reveals convergence across 400 million years of evolution." Communicative & Integrative Biology **4**(1): 20.
- Yamaguchi, M. and T. Demura (2010). "Transcriptional regulation of secondary wall formation controlled by NAC domain proteins." Plant Biotechnology **27**(3): 237-242.
- Yamamoto, E., N. G. Lewis, et al. (1987). "5-hydroxyferulic acid in *Zea mays* and *Hordeum vulgare* cell walls." Phytochemistry **26**(7): 1915-1916.

- Yang, B. and C. E. Wyman (2004). "Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose." Biotechnology and bioengineering **86**(1): 88-98.
- Ye, Z. H. (1997). "Association of caffeoyl coenzyme A 3-O-methyltransferase expression with lignifying tissues in several dicot plants." Plant physiology **445**(4): 1341.
- Ye, Z. H., R. E. Kneusel, et al. (1994). "An alternative methylation pathway in lignin biosynthesis in Zinnia." The Plant Cell Online **6**(10): 1427.
- Yokoyama, R. and K. Nishitani (2004). "Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and Arabidopsis." Plant and cell physiology **45**(9): 1111.
- Youn, B., R. Camacho, et al. (2006). "Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4." Org. Biomol. Chem. **4**(9): 1687-1697.
- Yuan, J. S., K. H. Tiller, et al. (2008). "Plants to power: bioenergy to fuel the future." Trends Plant Sci **43**(8): 421-429.
- Zhang, K., Q. Qian, et al. (2006). "GOLD HULL AND INTERNODE2 encodes a primarily multifunctional cinnamyl-alcohol dehydrogenase in rice." Plant physiology **440**(3): 972.
- Zhang, Y., T. Culhaoglu, et al. (2011). "Impact of Lignin Structure and Cell Wall Reticulation on Maize Cell Wall Degradability." Journal of agricultural and food chemistry.
- Zhao, Q. and R. A. Dixon (2011). "Transcriptional networks for lignin biosynthesis: more complex than we thought?" Trends in plant science.
- Zhao, Q., H. Wang, et al. (2010). "Syringyl lignin biosynthesis is directly regulated by a secondary cell wall master switch." Proceedings of the National Academy of Sciences **407**(32): 14496.
- Zhong, R. (1998). "Dual methylation pathways in lignin biosynthesis." The Plant Cell Online **40**(12): 2033.
- Zubieta, C., P. Kota, et al. (2002). "Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase." The Plant Cell Online **44**(6): 1265.
- USDE Biofuels Joint Roadmap, Biomass to Biofuels Report 2006  
<http://genomicscience.energy.gov/biofuels/b2bworkshop.shtml>

# ANNEXES

# MATERIEL ET METHODES

## 1. Protocoles de bioinformatique

### 1.1. Recherche de séquences

Les séquences d'intérêt ont été recherchées sur diverses bases de données :

- Brachypodium : dans un premier temps, la séquence 4X organisée en BACs était consultable sur Brachypodium.org (<http://www.brachypodium.org/>). Sur ce site, l'outil Brachyblast (<http://blast.brachypodium.org/>) permettait dès l'automne 2008 l'alignement de séquences nucléiques et protéiques sur le génome de Brachypodium.
- Après l'annotation de la séquence 8X par l'IBI (IBI, 2010), le site du MIPS (Munich Information Center for Protein Sequence) a été mis en place et disponible au public (<http://mips.helmholtz-muenchen.de/plant/brachypodium/>). Ce site contient de nombreux outils relatifs au génome de Brachypodium et à la traduction des gènes annotés. Les recherches de séquences sur ce site peuvent se faire par BLAST sur le génome, par recherche de domaines protéiques, par noms de contigs ou par élément génétique à partir du nom correspondant à la position sur le génome ou du nom de la famille du gène prédit à l'URL suivante :  
<http://mips.helmholtz-muenchen.de/plant/brachypodium/searchjsp/index.jsp>.
- La base de donnée TAIR (<http://www.arabidopsis.org/>) regroupe les données concernant le génome d'*Arabidopsis thaliana*.
- Les autres séquences, protéiques ou nucléotidiques ont été obtenues respectivement sur les bases de données protein (<http://www.ncbi.nlm.nih.gov/protein/>) et nucleotide (<http://www.ncbi.nlm.nih.gov/nucleotide/>) du NIH.



## 1.2. Formatage et traduction

Certains outils d'analyse ne supportant que les entrées en format FASTA, le format des séquences obtenues a été modifié automatiquement sur le site fntseq (<http://www.bioinformatics.org/JaMBW/1/2/>).

Si les séquences obtenues étaient nucléotidiques, elles ont été traduites par l'outil de traduction d'Expasy (<http://web.expasy.org/translate/>). Cet outil traduisant dans les 6 cadres de lecture et proposant tous les codons initiateurs et terminateurs présents dans la séquence d'entrée, il a pu servir à la reconstruction à partir de la séquence génomique de gènes dont les introns et les exons avaient mal été identifiés dans la séquence 4X en 2008.

## 1.3. Alignements de séquences

Les séquences obtenues ont été alignées par le programme Multalin (<http://multalin.toulouse.inra.fr/multalin/>) (Corpet, 1998). Les options d'alignement par défaut ont été utilisées pour les séquences protéiques et seule la matrice de comparaison a été changée dans le cas d'alignements de séquences nucléotidiques (Blossum62 et DNA respectivement - pour plus d'information [http://multalin.toulouse.inra.fr/multalin/help.html#symbol\\_comparison\\_table](http://multalin.toulouse.inra.fr/multalin/help.html#symbol_comparison_table)). Pour faciliter l'interprétation visuelle des alignements multiples contenant un grand nombre de séquences protéiques, les paramètres des niveaux de consensus ont été modifiés : High consensus value à 100% (valeur par défaut 90%) et Low consensus value à 80% (valeur par défaut 50%). Ainsi seuls les acides aminés conservés dans toutes les espèces analysées apparaissent dans la couleur de haut consensus, permettant l'identification de résidus de grande importance évolutive.

#### 1.4. Arbres phylogénétiques

Les séquences identifiées ont été entrées dans le mode « one click » du site phylogeny.fr développé par le laboratoire MAB (Méthodes et Algorithmes pour la Bio-informatique) du LIRMM (<http://www.phylogeny.fr/>) (Dereeper et al., 2008). Ce mode permet au programme de déterminer les paramètres par défauts à utiliser en fonction des données entrées. Seuls les paramètres d'affichage des résultats (type d'arbre phylogénétique, couleurs, organisation des branches et des nœuds...) sont ensuite modifiés pour faciliter l'interprétation visuelle de l'arbre phylogénétique de sortie et les relations entre groupes de protéines.

#### 1.5. Prédiction de structures protéiques 3D

Aucune protéine de *Brachypodium distachyon* n'ayant été cristallisée jusqu'à présent, seules des prédictions de nos protéines d'intérêt ont pu être réalisées. Les séquences de ces protéines ont d'abord été entrées dans le moteur de recherche du site PDBsum (<http://www.ebi.ac.uk/pdbsum/>) recensant les données structurales des protéines ayant été cristallisées. Ceci permet la détermination des protéines pouvant servir de matrice à la modélisation de nos protéines d'intérêt de par leur degré d'identité et d'autres facteurs tels que les espèces d'origines, leur fonction biologique ou les ligands inclus dans leurs complexes.

Cette étape préliminaire achevée, la séquence protéique d'intérêt est alors entrée sur le site du PBIL (Pôle BioInformatique Lyonnais) Geno3D ([http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d\\_automat.pl?page=/GENO3D/geno3d\\_home.html](http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html)). Les paramètres d'alignement par défaut sont conservés, mis à part « Filter query sequence » qui change les acides aminés sans importance pour la structure tertiaire en X (ou Xaa, acide aminé inconnu ou non spécifié). Les modifications de séquence qui pourraient nous intéresser pouvant potentiellement se situer à tous les niveaux dans la séquence, il convient de modifier ce paramètre sur « false » afin de ne

perdre aucune information de séquence. L'étape suivante consiste au choix de la ou des séquences matrices pour la modélisation, ou plus simplement de retrouver la ou les séquences choisies précédemment sur PDBsum. Cette dernière servira alors de matrice pour la modélisation par homologie de notre protéine d'intérêt. Les coordonnées tridimensionnelles de chaque atome de notre protéine seront alors compilées dans un fichier .pdb qui pourra être ouvert par des programmes de visualisation tels que Pymol ou SwissPDB viewer.

### 1.6. Visualisation et modification des protéines modélisées

Les prédictions obtenues par Geno3D sont ouvertes avec le programme Pymol version de 2006 (<http://www.pymol.org/>). Les modèles de protéines sont visualisées sous les formes de repliements de leur structure secondaire (« cartoon ») ou selon l'encombrement total de la molécule, qui permettront l'analyse des modifications causées par les mutations dans les gènes codant pour ces protéines. Les acides aminés d'intérêt particulier (sites actifs, sites de fixation au co-facteur, etc...) peuvent être colorés pour faciliter leur repérage et représentés sous forme de « sticks » pour évaluer leur encombrement par rapport à la structure complète.

Le « mutagenesis wizard » permet de modifier un acide aminé en un autre, correspondant aux mutations ponctuelles présentes dans les lignées de la collection de mutants de Versailles : sélectionner l'acide aminé à changer, choisir le nouvel acide aminé, « apply », « done ».

Ceci peut permettre d'évaluer l'impact des mutations dans les gènes mutés sur l'activité d'une protéine en identifiant des modifications susceptibles de modifier sa structure ou l'accès des ligands à leurs sites cibles.

## 1.7. Prédiction de l'impact des mutations sur l'activité d'une protéine d'intérêt

L'impact d'une mutation sur l'activité d'une protéine peut être évaluée par la conservation de l'acide aminé muté dans le règne végétal et son implication dans les sites actifs ou de fixation de ligands. Le site SIFT (Sorting Intolerant From Tolerant) (<http://sift.bii.a-star.edu.sg/>) est un outil de prédiction de l'impact d'une substitution d'acides-aminés, basé sur les homologies de séquences et les propriétés physiques des acides aminés. Le programme calcule des scores associés à chaque substitution possible. Une valeur de ce score inférieure à 0,05 suggère une substitution intolérante, c'est à dire qui aura un effet important sur l'activité de la protéine.

## 2. Protocoles de biologie moléculaire

### 2.1. Extraction d'ADNg au CTAB (Arabidopsis et Brachypodium, tous organes)

- Prélèvement des échantillons : les placer immédiatement dans la glace ou l'azote liquide et conserver à -20°C ou -80°C.

- Broyage :

à l'azote liquide avec un pilon et un mortier (recommandé pour les entrenœuds de Brachypodium), récupérer la poudre dans un tube et y ajouter 500 µL de tampon d'extraction et vortexer.

ou

au broyeur mécanique fastprep® de chez Krackeler Scientific (recommandé pour les tissus peu lignifiés) dans un tube contenant une bille en céramique, quelques billes en verre et 500 µL de tampon d'extraction.

- Isolation des acides nucléiques : incuber les échantillons 30 minutes à 60°C, ajouter 500 µL de Chloroforme:Alcool Isoamylique (24 : 1), vortexer, centrifuger 10 minutes à 4° à 5000 rpm puis récupérer la phase supérieure, la phase aqueuse contenant les acides nucléiques.
- Précipitation des acides nucléiques : ajouter 1 mL d’Ethanol absolu et 40 µL d’acétate de sodium (3M, pH5,2) et laisser précipiter à -20°C minimum 2h, puis centrifuger 30 minutes à 4°C à 13000 rpm, éliminer le surnageant, rincer le culot dans 1 mL d’Ethanol 70%, centrifuger à nouveau pendant 30 minutes puis éliminer le surnageant et laisser sécher le culot pour éliminer l’Ethanol résiduel.
- Reprise du culot d’acides nucléiques : le culot est repris dans 50 µL de TE-RNase (pH8), incubé 30 minutes à 37°C pour éliminer les ARN et ne conserver que les ADN.
- Le produit de l’extraction peut être dosé sur gel.

Tampon d’extraction :

- CTAB en poudre : 2%
- NaCl 1,4 M : 0,28 mol
- EDTA 20 mM : 100 mM
- Tris-HCl 100 mM : 1 M
- β-mercaptoéthanol : 0,2%

## 2.2. Extraction d’ARN (Arabidopsis et Brachypodium)

- Prélèvement des échantillons : les placer immédiatement dans la glace ou l’azote liquide et conserver à -20°C ou -80°C. Les échantillons sont systématiquement récoltés 4h après l’allumage des lumières des chambres de culture, correspondant au maximum d’expression de la majorité des gènes de la voie de biosynthèse des monolignols et permettant de limiter les variations d’expression dues au cycle circadien.

- Broyage à l'azote liquide dans un mortier. Les échantillons broyés peuvent être recongelés dans l'azote liquide et conservés à -80°C.

- Extraction avec le kit RNeasy Plant Mini Kit de Qiagen, selon les recommandations du fabricant :

1. Sous la sorbone, resuspendre le matériel végétal dans 450 µL de tampon RLT + β-mercaptoéthanol et incubé à 56°C pendant 3 minutes.
2. Transférer, sous la sorbone, les 450 µL dans la colonne VIOLETTE et centrifuger à 8000 g, 4°C pendant 2 minutes.
3. Récupérer, sous la sorbone, le surnageant et le mélanger dans un eppendorf à 1/2 volume d'éthanol 96%.
4. Transférer dans une colonne ROSE puis centrifuger à 8000 g, 4°C pendant 2 minutes et éliminer le produit d'éluat.
5. Option DNase sur colonne
  - a. Ajouter 350 µL de tampon RW1 à la colonne et laisser agir 1 minute avant de centrifuger à 8000 g, 4°C pendant 2 minutes.
  - b. Ajouter le mix contenant 10 µL de DNase Qiagen + 70 µL de tampon RDD, incubé 20 minutes à température ambiante.
  - c. Ajouter 350 µL de tampon RW1 à la colonne et laisser agir 1 minute avant de centrifuger à 8000 g, 4°C pendant 2 minutes. Passer directement à l'étape 6.
6. Ajouter 700 µL de tampon RW1 à la colonne et laisser agir 2 minutes avant de centrifuger à 8000 g, 4°C pendant 2 minutes.
7. Laver 2 X avec 500 µL de tampon RPE et centrifuger à 8000 g, 4°C pendant 1 minute en éliminant l'éluat.
8. Centrifuger une fois de plus à vide afin de bien sécher la colonne, à 8000 g, 4°C pendant 1 minute.

9. Enfin, l'élution est réalisée en deux temps, en déposant 2 X 20 µL d'eau RNase free et en centrifugeant à 8000 g, 4°C pendant 1 minute, après avoir laissé 5 minutes l'eau décrocher les molécules d'ARN de la colonne.

- Les ARN sont alors conservés à -80°C.

### 2.3. Rétrotranscription

- Dosage des ARN au Nanodrop (de Thermo scientific), la réaction se fera à partir de 1 ou 2 µg d'ARN dans 10 µL d'eau ou de TE.

- Rétrotranscription : à chaque échantillon, ajouter 1 µL de dNTP et 1 µL d'OligoDT et incuber 5 minutes à 65°C puis une minute sur glace pour hybrider les OligoDT sur la queue polyA des ARN messagers. Puis ajouter 4 µL de tampon, 1 µL de DTT et 1 µL de SSRTII et incuber 5 minutes à température ambiante puis 50 minutes à 42°C.

- La réaction est arrêtée à 70°C pendant 15 minutes.

- Contrôle : pour chaque échantillon, un mélange réactionnel sera préparé sans ajouter de rétrotranscriptase afin de vérifier l'absence d'ADN résiduels dans les ARN.

- Les ADNc obtenus peuvent être dosés au nanodrop pour vérifier que tous les échantillons sont à une concentration identique et seront conservés à -20°C.

### 2.4. PCR

Les ADN (génomiques ou complémentaires) sont amplifiés par PCR (Polymerase Chain Reaction) en fonction des utilisations qui seront faites des produits avec les mélanges réactionnels suivants :

- PCR pour génotypage ou quantification : avec *Taq* DNA Polymerase (Invitrogen), programme T dénaturation : 95°C, Tm 60°C.

H<sub>2</sub>O : 17,95 µL

Tampon : 2,5 µL

Oligonucléotide 5' : 1,3 µL

Oligonucléotide 3' : 1,3 µL

MgCl<sub>2</sub>: 0,75 µL

dNTP: 0,5 µL

Taq: 0,2 µL

Matrice: 0,5 µL

- PCR pour séquençage ou clonage : Phusion® High-Fidelity DNA polymerase (NEB), programme T dénaturation : 98°C, T<sub>m</sub> touchdown 65-55°C.

H<sub>2</sub>O : 32,5 µL

Tampon : 10 µL

Oligonucléotide 5' : 2 µL

Oligonucléotide 3' : 2 µL

dNTP: 1 µL

Taq: 0,5 µL

Matrice: 2 µL

## 2.5. Clonage GATEWAY® (Invitrogen)

- Production du fragment à cloner par RT-PCR par polymérase Phusion®, entre le codon initiateur et le codon STOP de chaque gène à cloner, avec des oligonucléotides contenant les séquences attA et attB s'ajoutant à leurs extrémités.
- Extraction sur gel: Les produits PCR sont mis à migrer sur gel d'agarose 1% (TAE) et chaque bande correspondant à un fragment amplifié est découpée et purifiée avec le kit



NucleoSpin® Extract II (Macherey-Nagel) selon les recommandations du fabricant afin de les débarrasser des résidus de réactifs de la PCR et éventuellement de bandes aspécifiques.

- Introduction du gène dans le vecteur donneur pDONR207 par recombinaison BP : 30 à 300 ng du produit à cloner ; 0,5 µL de vecteur ; 1 µL de tampon de réaction et TE qsp 8 µL laissés 2 minutes sur glace, puis ajout d'1 µL de BP clonase et incubation 60 minutes à 28°C et enfin 0,5 µL de protéinase K pour arrêter la réaction. La recombinaison a lieu entre les adaptateurs attA et attB situés de part et d'autre du gène à cloner, et le polylinker du vecteur donneur.

- Transformation des bactéries : 2 µL du produit BP est introduit dans un tube de 30 µL de bactéries électrocompétentes Top10 (Invitrogen) qui sont alors placées dans une cuve et électroporées à 1,25 V (200 Ω) pour faire pénétrer le plasmide. On y ajoute immédiatement 1 mL de LB liquide et on laisse les bactéries incuber 1h à 37°C sous agitation. On les transfère ensuite sur boîte de Pétri contenant du milieu LB + Gentamycine, le plasmide pDONR207 contenant un gène de résistance à cet antibiotique permettant la sélection de colonies l'ayant intégré, et elles sont mises à incuber à 37°C pendant une nuit. Les colonies résistantes sont alors transférées en milieu de culture liquide LB + Gentamycine pendant une nuit ou 24h sous agitation.

- Extraction du plasmide : les cultures liquides sont centrifugées (10 min à 5000 rpm) et le surnageant est éliminé. Les plasmides sont extraits par le protocole d'extraction décrit au paragraphe (2.6). Ces plasmides sont envoyés à séquencer afin de vérifier la bonne insertion du gène. Une simple vérification par PCR, éventuellement suivie d'une digestion par une enzyme de restriction sélectionnée pour l'identification de chaque insert, permet de continuer le clonage en attendant le retour des séquences. Il est important d'effectuer un glycérol stock à cette étape, le vecteur donneur permettant de transférer le gène d'intérêt dans divers plasmides de la collection Gateway®.

- Introduction du gène dans le vecteur de destination pGWB2 par recombinaison LR : 100 à 300 ng du vecteur d'entrée ; 2 µL de vecteur de destination ; 4 µL de tampon de réaction LR, TE qsp 16 µL et 4 µL de LR clonase sont incubés 60 minutes à 25°C puis un ajout de 2 µL de protéinase K est fait pour arrêter la réaction. La recombinaison a lieu entre les adaptateurs attL situés de part et d'autre du gène à cloner dans le vecteur d'entrée, et les séquences attR du vecteur de destination contenant dans ce cas précis un promoteur 35S de CaMV permettant l'expression du gène sous contrôle d'un promoteur constitutif.

- Transformation des bactéries : 2 µL du produit LR est introduit dans un tube de 30 µL de bactéries électrocompétentes Top10 (Invitrogen) qui sont alors placées dans une cuve et électroporées à 1,25 V (200 Ω) pour faire pénétrer le plasmide. On y ajoute immédiatement 1 mL de LB liquide et on laisse les bactéries incuber 1h à 37°C sous agitation. On les transfère ensuite sur boîte de Pétri LB + Kanamycine, le plasmide pGWB2 contenant un gène de résistance à cet antibiotique permettant la sélection de colonies l'ayant intégré, et elles sont mises à incuber à 37°C pendant une nuit. Les colonies résistantes sont alors transférées en milieu de culture liquide LB + Gentamycine pendant une nuit ou 24h sous agitation.

- Extraction du plasmide : protocole identique à celle de l'étape précédente, et seule une vérification par PCR est utile.

- Incorporation du plasmide dans *Agrobacterium tumefaciens* : 2 µL du plasmide extrait sont électroporés aux Agrobactéries électrocompétentes (souche C58pMP90) à 1,25 V (200 Ω). . On y ajoute immédiatement 1 mL de LB liquide et on laisse les bactéries incuber 1h à 28°C sous agitation, puis elles sont transférées sur boîtes de LB + Rifampicine + Gentamycine + Kanamycine (RGK) et incubées 24 à 48h à 28°C. Les colonies résistantes sont repiquées en milieu liquide LB + RGK.

## 2.6. Extraction plasmidique par lyse alcaline

- Une culture liquide dans un milieu LB + antibiotique est préparée la veille.
- Osmolyse des bactéries : transférer la culture bactérienne dans des tubes de 1,5 mL et centrifuger 2 minutes à 13000 rpm. Eliminer le surnageant et reprendre le culot dans 100 µL de Solution A et laisser 2 à 5 minutes à température ambiante.
- Dénaturation de l'ADN bactérien : ajouter 200 µL de Solution de soude, mélanger par retournements du tube et ne pas laisser agir plus de 5 minutes.
- Neutralisation : ajouter 150 µL de Solution B, mélanger par retournements du tube puis incubé 10 minutes sur la glace avant de vortexer. Cette étape permet de renaturer brutalement l'ADN bactérien. L'ADN génomique étant très long et ses brins n'étant pas restés accrochés ensemble lors de l'étape précédente, il n'aura pas le temps de se renaturer et restera insoluble. L'ADN plasmidique étant plus court et ses brins restant accrochés ensemble, il restera en solution.
- Récupération de l'ADN plasmidique : centrifuger 10 minutes à 13000 rpm à température ambiante et transférer le surnageant récupéré dans un tube eppendorf 1,5 mL.
- Précipitation à l'isopropanol : ajouter au surnageant 0,7 Vol ( $\approx 350$  µL) d'isopropanol, laisser agir 5 minutes à température ambiante puis centrifuger 5 minutes à 13000 rpm à température ambiante et éliminer le surnageant. Rincer le culot dans 1 mL d'éthanol 70%, vortexer brièvement et centrifuger 10 minutes à 13000 rpm à température ambiante. Eliminer à nouveau le surnageant et bien laisser sécher le culot pour éliminer l'éthanol résiduel.
- Reprise de l'ADN plasmidique : dans 30 µL d'eau ou de TE-RNase. Le plasmide peut alors être conservé à  $-20^{\circ}\text{C}$ .

Solution A (conservée à  $4^{\circ}\text{C}$ ):

- Tris-HCl, 1M (pH 8) 10 mL
- EDTA, 100 mM (pH 8) 20 mL

- RNAaseA, 0,01 mg.ml-1

- Glucose (M=180) 1,8 g

- H<sub>2</sub>O qsp 200 mL

Solution de soude :

- NaOH 0,2 M

- SDS 1%

Solution B (pH 4,8) :

- Acétate de K 59 g

- Ac.OH glacial 100% 23 mL

- H<sub>2</sub>O qsp 200 mL

- Tampon TE\*, pH8

- Tris-HCl EDTA, (TE pH8) 1,5 mL

- Rnase (30 ng/μL) 3 μL

## 2.7. Transformation d'Arabidopsis thaliana

- Les lignées d'Arabidopsis à transformer doivent avoir été semées environ 5 semaines avant la transformation pour être à un stade de floraison pas trop avancé (hampes florales d'environ 10 cm). Les plantes peuvent être préparées en coupant les siliques déjà formées, ce qui augmentera la proportion de graines transformées.

- Prévoir une culture LB + antibiotiques de 500 mL la veille de la transformation (peut avoir été initiée 2 jours plus tôt dans une culture de 50 mL). Idéalement la DO de la culture doit être  $\geq 0,8$ .

- Préparation des Agrobactéries : Centrifuger 10 minutes à 6500 RPM et 4°C dans 2 tubes « biberons » stériles (2 x 250 mL). Éliminer le surnageant et reprendre chaque culot dans 125 mL de milieu d'infiltration. Mélanger les deux suspensions.

- Transformation des plantes : tremper les organes floraux 2 minutes dans la suspension bactérienne pendant 2 minutes.

- Récolte des graines : les plantes transformées sont remises en culture jusqu'à leur séchage, leurs graines sont ensuite récoltées, séchées et stérilisées.

Milieu d'infiltration :

- 25 g de saccharose
- 100  $\mu$ L de Silwet
- H<sub>2</sub>O qsp 500 mL

## 2.8. Stérilisation de graines d'*Arabidopsis*

- Premier lavage : ajouter 800  $\mu$ L de « Seed Clean » à un fond de tube de graines, agiter pendant 10 minutes puis éliminer la solution.

- Second lavage : ajouter 800  $\mu$ L de « Seed Wash », agiter pendant 10 minutes puis éliminer la solution.

- Troisième lavage : idem que précédemment.

- Séchage : retirer à la pipette autant que possible le reste de Seed Wash puis laisser sécher sous la hotte pendant minimum 1 heure.

Seed Clean :

- Ethanol 70%
- Triton 0,01%
- H<sub>2</sub>O qsp 100 mL

Seed Wash :

- Ethanol 95%

## 2.9. Sélections des transformants primaires (*Arabidopsis* T1)

- Semis *in vitro* : semer à la volée une centaine de graines stériles provenant des plantes transformées, sur boîtes de pétri de Milieu *Arabidopsis* + Céfotaxime (inhibe la croissance des Agrobactéries) + antibiotique correspondant à la résistance contenue dans le vecteur inséré (ici pGWB2 : Kanamycine 50 µg/mL), vernaliser 48h à 4°C à l'obscurité puis disposer les boîtes en phytotron ou dans les tours de culture pendant 10 jours.
- Repiquage sur terreau : Les plantules résistantes seront vertes et présenteront leurs 2 cotylédons et éventuellement 2 feuilles. Elles seront repiquées en serre en les déposant sur du terreau et en y introduisant délicatement les racines.

Milieu Arabidopsis :

- Milieu Arabidopsis Duchefa 6g
- Agar Biomar 3,5 g ou Kalys 3 g
- Ca(NO<sub>3</sub>)<sub>2</sub> 1 M 700 µL
- H<sub>2</sub>O ultra-pure qsp 450 mL
- Ajuster le pH entre 5,6 et 5,7 (au KOH),
- H<sub>2</sub>O ultra-pure qsp 500mL
- Facultatif : Saccharose 5 g

## 2.10. Sélection des transformants secondaires (*Arabidopsis* T2)

- Semis *in vitro* : semer à la volée une centaine de graines stériles provenant des transformants primaires sur boîtes de pétri de Milieu *Arabidopsis* + antibiotique correspondant à la résistance contenue dans le vecteur inséré (ici pGWB2 : Kanamycine 50 µg/mL), vernaliser

48h à 4°C à l'obscurité puis disposer les boîtes en phytotron ou dans les tours de culture pendant 10 jours.

- Comptage de la ségrégation : la descendance des lignées de transformants primaires n'ayant intégré qu'une seule copie du transgène présenteront une ségrégation de type mendelienne avec  $\frac{1}{4}$  de leur descendance ne contenant aucune copie du transgène,  $\frac{1}{2}$  de leur descendance en contenant une seule et  $\frac{1}{4}$  de leur descendance en contenant 2. Les lignées contenant une du transgène seront sélectionnées et pousseront sur le milieu de sélection, ce qui donnera au total  $\frac{1}{4}$  de plantes sensibles et  $\frac{3}{4}$  de plantes résistantes.

- Repiquage sur terreau : Les plantules résistantes des lignées ségrégeant selon le modèle  $\frac{3}{4}$   $\frac{1}{4}$  seront repiquées en serre.

#### 2.11. Sélection de transformants tertiaires (*Arabidopsis* T3)

- Semis *in vitro* : semer à la volée une centaine de graines stériles provenant des transformants secondaires sur boîtes de pétri de Milieu *Arabidopsis* + antibiotique correspondant à la résistance contenue dans le vecteur inséré (ici pGWB2 : Kanamycine 50  $\mu\text{g}/\text{mL}$ ), vernaliser 48h à 4°C à l'obscurité puis disposer les boîtes en phytotron ou dans les tours de culture pendant 10 jours.

- Comptage de la ségrégation : la descendance des lignées de transformants secondaires portant une copie du transgène à l'état homozygote sera totalement résistante au milieu de sélection.

- Repiquage sur terreau : Les plantules résistantes des lignées 100% résistantes seront repiquées en serre.

### 3. Protocoles de cytologie

#### 3.1. Coupes transversales au vibratome

- Réhydratation et inclusion des organes : Les organes à couper sont placés dans l'eau la veille si nécessaire. Ils sont ensuite inclus dans de l'agarose à 8% dissoute dans de l'eau puis laissés à reposer jusqu'à polymérisation.
- Ces inclusions sont placées sur la plaque du vibratome (Leica VT1000S) et coupées par des lames de rasoir (speed 3, freq 7) à une épaisseur de 50 à 80  $\mu\text{m}$ .
- Elles sont récupérées dans de l'eau et peuvent être conservées plusieurs jours avant observation. Si elles doivent être conservées pour une durée plus longue, elles pourront être conservées dans de l'Ethanol à 70% et réhydratées avant d'être observées.

#### 3.2. Coloration de lignines au phloroglucinol (réactif de Wiesner)

- Des coupes transversales de chaque échantillon et leurs témoins sont déposées sur des lames et l'eau dans laquelle elles étaient en suspension est éliminée.
- Quelques gouttes de Phloroglucinol (50mg/mL) sont déposées sur les coupes, pour une durée de 30 secondes à 3 minutes. La quantité et la durée de la coloration peuvent varier en fonction de l'état de lignification des échantillons (pour des échantillons très lignifiés, il n'y aura besoin que de 1 à 2 gouttes de réactif pendant quelques dizaines de secondes) mais pour pouvoir être comparables, les échantillons et leurs témoins doivent systématiquement subir le même traitement.
- Le réactif est éliminé et remplacé soit par de l'eau pour une observation immédiate, soit par un mélange Ethanol : HCl (9 :1) si un délai est prévu avant leur observation.

Cette coloration est spécifique des cinnamaldehydes et colore en rouge/rose les groupements coniferaldehydes terminaux et en rouge/violet les groupements sinapaldehydes terminaux



(Kim et al., 2002, voir figure). La coloration rouge est considérée comme étant représentative de la quantité totale de lignines mais dans certains cas les groupements cinnamaldéhydes ne sont pas représentatifs de la quantité globale de lignines.



Photographies de solutions de vanilline, d'hydroxycinnamaldehydes 1G et 1S et de deux hydroxycinnamaldehydes couplés en 8-O-4 3GS et 3SG traitées au phloroglucinol. Les hydroxycinnamaldehydes sont les composants les plus marqués (Kim et al. 2002).

### 3.3. Coloration de Maüle

- Des coupes transversales de chaque échantillon et leurs témoins sont déposées sur des lames et l'eau dans laquelle elles étaient en suspension est éliminée. Les colorations doivent systématiquement être faites en parallèle sur les échantillons et leurs témoins (mêmes quantités de réactifs et mêmes durées).
- Quelques gouttes de  $\text{KMnO}_4$  (1%) sont déposées sur chaque coupe et laissées 5 à 7 minutes, puis le  $\text{KMnO}_4$  est éliminé.
- Les coupes sont rincées à l'eau 1 minute, puis l'eau est éliminée.
- Les coupes sont alors décolorées par une goutte de  $\text{HCl}$  (30%) pendant 1 minute, puis l'acide est éliminé.
- Les coupes sont rincées à l'eau 1 minute, puis l'eau est éliminée.
- Quelques gouttes de  $\text{NaHCO}_3$  (5%) sont déposées sur les coupes, recouvrir d'une lamelle et observer rapidement.

Cette coloration permet de colorer qualitativement les unités syringiles des lignines en rouge-orangé et les unités coniféryles en marron.

### 3.4. Hybridations *in situ*

Les sondes sens et anti-sens sont produites à partir de produits PCR réalisés avec une amorce spécifique du gène et une amorce « hybride » qui est formée d'une séquence spécifique du gène et d'une séquence reconnue par la reverse transcriptase T7 (TGTAATACGACTCACTATAGGGC).

Après purification, ces produits serviront de matrice pour la transcription *in vitro* des sondes à l'aide du kit PROMEGA, selon les instructions du fabricant et avec des DIG-UTP .

Les sondes sont alors précipitées avec 20 ml d'acétate d'ammonium 10 M et 240 ml d'éthanol 100% froid pendant 1 nuit à -20°C. Après une centrifugation à 15000 g à 4°C pendant 30 min, le culot est rincé avec de l'éthanol 70% froid, puis séché et repris dans 40ml d'eau-BIOFIT ou TE stérile si la sonde à une longueur inférieure à 500 nucléotides ou dans 50 ml de tampon carbonate si sa taille est supérieure à 500 nucléotides. Après avoir repris les sondes dans 50 ml de tampon carbonate (Na<sub>2</sub>CO<sub>3</sub> 120 mM, NaHCO<sub>3</sub> 80 mM pH 10,2), une incubation est réalisée à 60°C. Le temps d'incubation est déterminé grâce à la formule suivante :  $t = (L_i - L_f) / (K * L_i * L_f)$  où t : durée de l'incubation en minutes ; L<sub>i</sub> : longueur initiale de la sonde en kb ; L<sub>f</sub> : longueur finale de la sonde = 0,15 kb ; K : constante = 0,11 kb/min. La réaction est arrêtée par 10 ml acide acétique (10%), 12 ml acétate de sodium (3 M pH 4,8) et 312 ml éthanol 100% froid. Puis les sondes sont précipitées à - 80°C au moins 30 min. Après centrifugation le culot est lavé avec de l'éthanol 70% froid. Puis la sonde est reprise dans 40 ml d'eau BIOFIT ou 20 ml d'eau BIOFIT + 20 ml de formamide. Enfin, la qualité des ARN synthétisés est vérifiée par Dot Blot.

Les tissus végétaux sont fixés dans du paraformaldéhyde sous vide pendant 60 minutes dans la glace, puis une nuit à 4°C. Ils sont ensuite déshydratés par des bains successifs de dilutions d'éthanol, colorés à l'éosine puis lavés dans des bains d'histoclear® avant d'être inclus dans la paraffine pour pouvoir être coupés.

Les coupes de 8  $\mu\text{m}$  d'épaisseur sont effectuées sur un microtome Leica RM2025 puis posées sur des lames chargées positivement. Celles-ci sont alors traitées à la protéinase K, facilitant l'accès aux ARN cibles, lavées dans des bains successifs d'éthanol et d'histoclear®, puis traitées à l'anhydride acétique dans du tréthanolamine, évitant les hybridations aspécifiques. Elles sont ensuite déshydratées dans des solutions successives d'éthanol.

Les lames sont préhybridées 1 h à la température d'hybridation dans 10mL de tampon de préhybridation (5 ml de formamide désionisée ; 2.5 ml de 20X SSC (concentration finale 5X) ; 100  $\mu\text{l}$  d'héparine stock (concentration finale 50  $\mu\text{g}/\text{ml}$ ) ; 100  $\mu\text{l}$  d'ARNt Stock (concentration finale 100  $\mu\text{g}/\text{ml}$ ) ; 10  $\mu\text{l}$  de tween 20% (0.1%)). Elles sont ensuite hybridées par incubation durant une nuit à la température d'hybridation avec la sonde ARN "DIG". Pour une lame, 1 à 9 ml de sonde sont dilués dans 9 à 1 ml d'eau BIOFIT suivant la concentration de la sonde. Après une dénaturation de 2 min à 80°C, on ajoute 210 ml de tampon DAKO. Les lames sont alors mises à hybrider dans une boîte humide (mélange 50% formamide, 50% 2 SSC sur du papier propre). Des lavages permettent de dissocier les hybrides aspécifiques.

La détection des hybrides ARN-ARN marqués avec la digoxigénine utilise des fragments Fab anti-digoxigénine couplés à la phosphatase alcaline. Pour cela les coupes sont d'abord préincubées dans un tampon contenant du "blocking reagent" puis dans un tampon contenant de la BSA (albumine de sérum bovin, Sigma) pour bloquer les sites non-spécifiques. Après incubation avec l'anticorps, l'excès d'anticorps est éliminé par des lavages en présence de BSA et d'un détergent le Triton.

La réaction de coloration est initiée à pH alcalin. Au tampon 4 on ajoute les deux substrats: le X-P ou 5-bromo-4-chloro-3-indolyl-phosphate et le NBT ou "Nitroblue tetrazolium chloride". Un précipité bleu se forme dans un laps de temps qui varie de 1 h à 2 jours en fonction de la quantité d'hybrides, donc en fonction de l'abondance des ARNm cibles. La réaction de coloration est basée sur l'oxydation du X-P et sur la réduction couplée du NBT qui donnent

toutes les deux un précipité bleu. Ce précipité est insoluble dans l'eau et adhère aux coupes. Les lames sont ensuite rincées à l'eau puis montées avec du citifluor entre lame et lamelle et observées au microscope.

#### 4. Protocoles d'analyses de lignines

##### 4.1. Protocole de thioacidolyse

Les thioacidolyses sont réalisées sur environ exactement 15 mg d'échantillon broyé et sec, mis en présence d'étalon interne (C21 0,5 mg). Les échantillons traités sont soit des tiges broyées et sèches non extraites, soit du résidu pariétal (RP) obtenu à partir de ces tiges et dont on connaît le taux de Lignines Klason. Le protocole employé est décrit dans les articles suivant (Lapierre *et al.*, 1995, 1999). Les résultats sont exprimés en  $\mu$ moles de monomères H, G et S par gramme d'échantillon ou par gramme de LK.

#### 5. Protocoles de biochimie

##### 5.1. Extraction et dosage des protéines totales

- Extraction : Broyer quelques mg de tissus végétaux en poudre fine dans l'azote liquide, transvaser la poudre dans un tube Eppendorf et ajouter 0,5 à 1 mL de tampon d'extraction des protéines. Agiter vigoureusement le tube puis centrifuger 15 min à 14000 rpm et 4°C pour éliminer les débris végétaux puis reprendre le surnageant dans un nouveau tube.

- Dosage : La quantité totale de protéines est obtenue par comparaison avec une gamme étalon de BSA. Les dosages s'effectuent sur une plaque de micro-titration à l'aide d'un spectrophotomètre lecteur de microplaques à 595nm. La gamme étalon se fait avec de 0 à 8  $\mu$ g de BSA mélangée à 200 $\mu$ L de Bradford, de 1 à 10  $\mu$ L de chaque échantillon est mélangé à

200µL de réactif de Bradford et 4 blanc sont faits avec 200µL de Bradford sans protéines. La coloration est instantanée et reste stable une quinzaine de minutes.

Tampon d'extraction :

- Tris-HCl 0,1M pH7,5
- PEG 6000 à 0,5%
- PolyclarAT ou PVP à 0,4%
- β-mercaptoethanol 15mM à 0,1%

## 5.2. Test d'activité enzymatique CAD

La cinétique s'effectue dans une microplaque de type ELISA sur le lecteur de micro plaques par une mesure à 405nm toutes les 30 secondes pendant 1h pour mesurer l'apparition de l'aldéhyde (jaune). Cette cinétique utilise la capacité de la CAD à catalyser la réaction de l'alcool coniférylique en coniféraldéhyde, bien que dans la voie de biosynthèse des monolignols elle réalise la réaction opposée :

- Contrôle en l'absence de co-facteur : 5 à 20 µg d'extrait protéique total + tampon Tris 0,1M pH8,8 qsp 300µL + 50µL d'alcool coniférylique/sinapylique 10mM.
- Contrôle de l'oxydation spontanée du substrat (sans protéines) : Tampon Tris 0,1M pH8,8 300 µL+ 50µL d'alcool coniférylique/sinapylique 10mM + 25µl de NAPD 10mM.
- Mesure d'activité enzymatique : 5 à 20 µg d'extrait protéique total + tampon Tris 0,1M pH8,8 qsp 300µL + 50µL d'alcool coniférylique/sinapylique 10mM + 25µl de NAPD 10mM.

# PREDICTIONS SIFT

