



UNIVERSITÉ D'ORLÉANS



ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT

INRA Unité de Recherche Amélioration, Génétique et Physiologie Forestières
(UR 0588)

THÈSE présentée par :
Wassim LAKHAL

Soutenue le : **18 décembre 2013**

pour obtenir le grade de : **Docteur de l'université d'Orléans**
Discipline/ Spécialité : Biologie moléculaire et cellulaire

**Etude fonctionnelle de trois facteurs de
transcription impliqués dans la formation de
la paroi secondaire chez le peuplier**

THÈSE dirigée par :

Annabelle DEJARDIN

Chargée de recherche, INRA

RAPPORTEURS :

Jacqueline GRIMA-PETTENATI

Jean-Louis JULIEN

Directrice de recherche, CNRS
Professeur, Université Blaise Pascal
(Clermont-Ferrand II)

JURY

Stéphane MAURY

Jacqueline GRIMA-PETTENATI

Jean-Louis JULIEN

Professeur, Université d'Orléans
Directrice de recherche, CNRS
Professeur, Université Blaise Pascal
(Clermont-Ferrand II)

Lise JOUANIN

Vincent COURDAVAULT

Directrice de recherche, CNRS
Maître de Conférence
Université François-Rabelais (Tours)
Chargée de recherche, INRA

Annabelle DEJARDIN

Remerciements

Ce manuscrit de thèse est le fruit du travail de tout un ensemble de personnes qui ont volontairement et activement participé à sa réalisation pour que je réussisse à l'accomplir et que j'obtienne le diplôme de doctorat. Je présente toutes mes excuses pour les personnes que j'aurai oublié de remercier par inattention.

*Tout d'abord, je remercie Mr **Gilles PILATE** qui m'a accueilli au sein de l'Unité Amélioration, Génétique et Physiologie Forestières. Je le remercie également pour avoir participé à l'encadrement de mon travail durant la troisième année de thèse mais aussi pour ses conseils précieux et également sa participation à la relecture du manuscrit de thèse.*

*Je remercie les membres du jury : Mesdames **Jacqueline GRIMA-PETTENATI**, **Lise JOUANIN** et **Annabelle DEJARDIN**, Messieurs : **Jean-Louis JULIEN**, **Vincent COURDAVAULT** et **Stéphane MAURY** d'avoir accepté de juger ce travail de thèse.*

*Je tiens à remercier le Département **EFPA**, la **Région Centre**, le projet **AMC2B** et le projet **TreeForJoules** géré par **Jean Charles LEPLE** pour le financement de cette thèse.*

*Je remercie vivement les membres de mon comité de thèse Mesdames **Nathalie LEBLANC-FOURNIER**, **Martine DECOVILLE** et Monsieur **Vincent COURDAVAULT** qui ont eu un rôle majeur dans l'orientation de mes travaux de thèse vers le bon chemin. Je remercie, de même, toutes les personnes de notre unité qui ont participé aux deux réunions de comité de thèse.*

*Je remercie également la **plateforme de séquençage de Strasbourg (Biopuces et Séquençage)**, le **Centre de Biophysique Moléculaire** du CNRS d'Orléans (**Martine DECOVILLE**), le laboratoire **Biomolécules et Biotechnologies Végétales (BBV)** de l'Université de Tours (**Audrey OUDIN** et **Vincent COURDAVAULT**) pour leur collaboration en étroite liaison avec mes travaux de thèse.*

*Je remercie très spécialement et avec beaucoup de chaleur et d'affection madame **Annabelle DEJARDIN** ma directrice de thèse pour m'avoir encadré durant ces trois années. Merci Annabelle pour m'avoir aidé à maîtriser le métier de la recherche. Merci pour avoir donné beaucoup de ton temps et de ton effort pour mener à bien ma thèse. Merci pour tous les*

sacrifices que tu as donnés en faveur de cette thèse. Je t'exprime tous mes sentiments de gratitude et de reconnaissance pour avoir corrigé mon manuscrit de thèse. Un grand merci également pour ta compréhension et ton caractère très doux.

*Mes sincères remerciements s'adressent aussi à Mesdames **Marie-Claude LESAGE-DESCAUSES**, **Nathalie BOIZOT**, **Odile ROGIER**, **Nadège MILLET** (merci pour tous les échanges amicaux et pour ta gentillesse exceptionnelle), **Françoise LAURANS** et **Véronique LAINE** (merci pour les discussions amicales et bravo pour ton intégration très rapide dans l'unité) pour leur participation très active et précieuse dans mes travaux de thèse, une participation qui a rajouté plus de valeur à mon manuscrit de thèse, mais également merci pour m'avoir appris des nouvelles techniques. Merci également à Mademoiselle **Pascale FERRIGNO** pour sa participation significative à la mise au point de la technique ChIP-SEQ sur peuplier (en collaboration étroite avec **Nathalie**). Je remercie aussi Madame **Vanina GUERIN** pour son aide pour l'extraction d'ADN, Madame **Véronique JORGE** pour m'avoir fourni du matériel végétal pour ma thèse et Mademoiselle **Claire LEMETTE** pour son aide au laboratoire de biologie cellulaire.*

*Je tiens à remercier aussi le Plateau Génoboïs, **Kevin ADER** et **Jean Paul CHARPENTIER** pour la participation aux analyses de lignines et cellulose.*

*Merci encore à **Marie-Claude**, **Odile**, **Nathalie**, **Françoise**, **Jean Paul** et **Kevin** pour la relecture de la partie matériel et méthodes de ce manuscrit.*

*Je remercie vivement l'Unité Expérimentale, **Christophe BOREL**, **Patrick POURSAT**, **José ALMEIDA**, **Ludovic PASQUIER**, **Benjamin DIMOURO**, **Jonathan RIAN**, **Jacky DESPRAS** pour toute leur aide et tous les travaux en serre.*

*Mes sincères remerciements à tous les chercheurs, ingénieurs, techniciens et personnel administratif de l'unité AGPF, **Vincent SEGURA** (merci pour le journal club et pour le CV en anglais), **Catherine** et **Jean Charles BASTIEN**, **Christian BRETON**, **Marc VILLAR**, **Philippe ROZENBERG**, **Nassim BELMOKHTAR**, **Corinne BURET**, **Frank ROGEON** (merci pour ton aide documentaire), **Karine ROBINEAUD**, **Jean Léandre HATON**, **Brigitte VIGUIER**, **Patricia MONTES**, **Véronique FREJAVILLE** et les agents d'accueil **Sandrine**, **Cyrielle** et **Christel** (merci pour ta gentillesse exceptionnelle).*

Un très grand merci à toutes les personnes qui ont participé à me faire l'agréable surprise de donner un cadeau à mon petit Abdou.

*Je remercie également tous les doctorants que j'ai côtoyés durant ma thèse, **Redouane, Amra, Fernanda, Alexandre, Maxime, Sylvain, Nicolas, Justine, Julien, Anne-Sophie** pour tous les bons moments de discussions et plaisanteries.*

*Je tiens à remercier les post-doctorants de l'Unité **Nicolas, Delphine, Jonathan** ainsi que les stagiaires et CDD **Brice, Baptiste, Marjolaine, Emna, Elise, Audrey, Vanessa, Flora**.*

*Je remercie avec un grand amour mon aimable épouse **Wefa** pour son soutien sans limites, son encouragement continu et sa sagesse exceptionnelle. Sans tes conseils et ton aide, je ne serais pas arrivé à ce niveau. Je m'excuse de ne pas être assez disponible pour notre petite famille durant toute la thèse car j'étais souvent face au PC ou plongé dans un article surtout durant la période de rédaction. Ton aide m'était indispensable pour la rédaction. Merci pour ta compréhension et tes sacrifices multiples et inimaginables pour le bien de ma thèse mais aussi de notre petite famille. Que chaque personne qui lira mon manuscrit sache qu'une grande histoire d'amour se cache derrière chaque mot et chaque lettre écrits dans ce manuscrit.*

*Merci à mon fils **Abd Arrahmen** (ou plus facilement Abdou) pour son aide précieuse durant la période de rédaction. Merci pour les mots que tu as tapés dans cette thèse et qui ont dû être supprimés (par confidentialité scientifique!).*

*Je remercie, avec une grande émotion et tendresse, mes parents **Jamila et Mouldi** pour leurs soutien sans ressemblant. Merci de m'avoir fourni tout ce qu'il faut, moralement et matériellement pour que j'arrive à ce stade d'étude. Merci pour vos sacrifices. Sans exagération, vous êtes parmi les meilleurs parents du monde.*

*Merci à mon frère **Karim** (a-ou-i), mes sœurs **Amani et Malak**, mes grands-parents **Chebil, Elhedi, Emnée, et Ftéma** (qui n'est plus parmi nous), ma belle-famille et surtout ma belle-mère **Najet**. Merci à tous mes oncles et tantes, mes amis en France, en Tunisie et ailleurs pour leur soutien.*

Merci aux personnes que j'aurais pu oublier de remercier, malgré mon grand effort, pour leur compréhension et leur pardon.

Liste des principales abréviations	1
Préambule	3
<u>INTRODUCTION ET OBJECTIFS DE THESE</u>	4
1- Le peuplier	4
1.1- Le genre <i>Populus</i>	4
1.2- Une essence d'intérêt économique	5
1.3- Un arbre modèle en biologie	6
2. Formation du bois et des parois secondaires	8
2.1- Anatomie du bois et fonctions pour l'arbre	8
2.2- Les différentes étapes de la formation du bois	9
2.3. Propriétés et dépôt des parois secondaires	9
3- Acteurs moléculaires connus de la formation du bois et des parois secondaires	11
3.1- Enzymes et gènes de structure	11
3.1.1- Gènes de la voie de biosynthèse des lignines	12
3.1.2- Gènes de la voie de biosynthèse de cellulose et des hémicelluloses	13
3.2- Facteurs de transcription impliqués dans la formation du bois et de la paroi secondaire	14
3.2.1- Facteurs de transcription à MADS-box	14
3.2.2- Facteurs de transcription LBD	15
3.2.3- Facteurs de transcription de la famille MYB	15
3.2.4- Facteurs de transcription de type NAC	18
3.2.5- facteurs de transcription de type KNOTTED-like (KNOX)	20
3.2.6- Facteurs de transcription de type WRKY	21
3.2.7- Facteurs de transcription de type HD-ZIP classe III	21
3.2.8- Facteur de transcription SHINE	21
4- Objectifs de la thèse	22
4.1- Choix des facteurs de transcription	24
4.2- Identification des cibles ADN des FT à l'échelle du génome	24
4.3- Etude fonctionnelle des FT dans des peupliers transgéniques	25
<u>MATERIEL ET METHODES</u>	27
1- Matériel végétal et échantillonnage	27
1.1- Plantes destinées à la ChIP	27
1.2- Lignées transgéniques	27
2- Extraction d'ARNs totaux et synthèse d'ADNc	28
3- Extraction d'ADN génomique	28
4- PCR	28

5- Q-RT-PCR	29
6- Production des plantes transgéniques	29
6.1- Réalisation des constructions	29
6.2- Transformation génétique	30
6.3- Caractérisation moléculaire des plantes transgéniques	31
7 - Obtention des anticorps polyclonaux anti- MYB90 et anti-MYB221	32
8 - Localisation cellulaire des FT en système cellulaire hétérologue	32
9- ChIP-Seq et ChIP-PCR	33
9.1- Pontage de la chromatine <i>in vivo</i>	33
9.2- Extraction et sonication de la chromatine	33
9.3- ChIP-Seq et ChIP-PCR	34
9.4- Séquençage et analyses bioinformatiques de l'ADN immunoprécipité	34
10- Etudes histologiques	36
10.1- Coloration Safranine-Bleu Alcian	36
10.2- Suivi de l'autofluorescence des parois en microscopie confocale	36
10.3-Détection par immunofluorescence des arabinoxylyanes	37
11- Dosage des lignines et de la cellulose	38
11.1- Extraction	38
11.2- Dosage des lignines	38
11.3- Dosage de la cellulose	39
 CHAPITRE 1	 41
<u>CHOIX DES FACTEURS DE TRANSCRIPTION</u>	
1- Etude d'expression	42
1.1- MYB090	42
1.2- MYB156 et MYB221	44
2- Clonage et analyse des séquences protéiques	45
2.1- MYB090	45
2.2- MYB156 et MYB221	46
3- Production d'anticorps et étude protéique	46
4- Localisation subcellulaire des FT	47
5- Conclusions	47

CHAPITRE 2 **49**

IDENTIFICATION DES CIBLES ADN DES FACTEURS DE TRANSCRIPTION A L'ECHELLE DU GENOME

1- Principes de l'immunoprécipitation de la chromatine	49
2- Détermination des cibles de MYB221 et MYB090 à l'échelle du génome	50
2.1- Mise en œuvre de la technique sur notre matériel biologique	51
2.1.1- Pontage de la chromatine <i>in vivo</i>	51
2.1.2- Sonication de la chromatine pontée	51
2.1.3- Immunoprécipitation de la chromatine	52
2.2- Résultats obtenus pour MYB221	53
2.2.1- Résultats de séquençage et d'analyses primaires pour MYB221	53
2.2.2- Classification fonctionnelle des cibles potentielles de MYB221	54
2.2.3- Recherche de motifs consensus dans les séquences des pics pour MYB221	54
2.2.4- Validation de 5 cibles de MYB221 par ChIP-qPCR	56
2.3- Résultats obtenus pour MYB090	57
2.3.1- Résultats de séquençage et d'analyses primaires pour MYB090	57
2.3.2- Classification des cibles de MYB090	57
2.3.3- Recherche de motifs consensus dans les séquences des pics pour MYB090	58
3- Discussion	59

CHAPITRE 3 **63**

FONCTION *IN PLANTA* DES FACTEURS DE TRANSCRIPTION

1- Production des lignées transgéniques	63
2- Sélection des lignées d'intérêt	65
3- Caractérisation phénotypique et moléculaire des plantes transgéniques	65
3.1- Peupliers MYB221 OE et MYB221-SRDX OE	66
3.1.1- Caractéristiques de croissance	66
3.1.2- Caractéristiques histologiques	67
3.1.3- Analyses biochimiques et de biomasse	69
3.1.4- Expression de gènes de biosynthèse des lignines, cellulose et xylanes	70
3.1.5- Expression de gènes cibles identifiés par ChIP-Seq	70
3.2- Peupliers MYB156 OE et MYB156-SRDX OE	71
3.2.1- Caractéristiques de croissance	71
3.2.2- Caractéristiques histologiques	74
3.2.3- Analyses biochimiques et de biomasse	74
3.2.4- Expression de gènes de biosynthèse des lignines, cellulose et xylanes	75
3.3- Peupliers MYB090 OE et MYB090-SRDX OE	75
3.3.1- Caractéristiques de croissance	75
3.3.2- Caractéristiques histologiques	77

3.3.3- Analyses biochimiques et de biomasse	79
3.3.4- Expression de gènes de biosynthèse des lignines, cellulose et xylanes	79
3.3.5- Expression de gènes cibles identifiés par ChIP-Seq	81
4- Discussion	81
 <u>DISCUSSION GENERALE</u>	 87
 <u>REFERENCES BIBLIOGRAPHIQUES</u>	 96
 <u>ANNEXES</u>	 105
Annexe 1	105
Annexe 2	106
Annexe 3	107
Annexe 4	108
Annexe 5	109
Annexe 6	111
Annexe 7	112
Annexe 8	113
Annexe 9	115

Liste des principales abréviations

4CL : hydroxycinnamate-CoA-ligase
ADN : acide désoxyribonucléique
ADNc : acide désoxyribonucléique complémentaire
AGP : protéine à arabinogalactanes
AMF : angle des microfibrilles de cellulose
ARN : acide ribonucléique
At : *Arabidopsis thaliana*
C3H : 4-coumarate-3-hydroxylase
C4H : cinnamate-4-hydroxylase
CAD : alcool cinnamylique déshydrogénase
CAZymes : carbohydate active enzymes
CCoAOMT : caffeoyl-CoA 3-O-méthyltransférase
CCR : cinnamoyl CoA réductase
CESA : cellulose synthase, sous-unité catalytique
CFP : cyan fluorescent protein
ChIP : immunoprécipitation de chromatine
ChIP-SEQ : immunoprécipitation de chromatine, couplée à un séquençage haut-débit
COMT : acide caféique O-méthyltransférase
CSL : cellulose synthase-like
Ct : Cycle Threshold
EAR : Ethylene-responsive element binding factor-associated Amphiphilic Repression
Eg : *Eucalyptus grandis*
EST : expressed sequence tag
EW : bois de printemps
F5H : ferulate hydroxylase
FLA : protéine à arabinogalactanes à domaine fascicline
FT : facteur de transcription
GT : glycosyltransférase
IRX : irregular xylem
Kb : kilo base
LW : bois d'été

Mb : mega base
MFs : microfibrilles
MYB : **myeloblastose**
OE : surexprimant (over expressor)
OW : bois opposé
Pb : paire de base
PAL : phenylalanine ammonia-lyase
PCR : réaction en chaîne par polymérase
PME : pectin methyl esterase
Pta : *Populus tremula* x *alba*
Ptr : *Populus trichocarpa*
Ptt : *Populus tremula* x *tremuloides*
Q-PCR : PCR quantitative
Qr : expression quantitative relative
RT-PCR : PCR reverse transcriptase
SMRE : Secondary wall MYB-Responsive Element
SRDX : domaine de répression dominante
TW : bois de tension
XTH : xyloglucan endotransglucosylase
YFP : Yellow Fluorescent Protein

Préambule

Ce travail de thèse a été réalisé au sein de l'Unité de Recherche Amélioration, Génétique et Physiologie Forestières, INRA Val de Loire, Centre d'Orléans. Il s'inscrit dans un des thèmes de recherche développés par l'Unité, à savoir le décryptage des mécanismes moléculaires de la formation du bois, chez le Peuplier, espèce modèle en biologie forestière, mais également espèce d'intérêt économique.

Mon manuscrit de thèse est organisé en 6 parties. Tout d'abord, dans une partie bibliographique introductive, je présenterai le peuplier et ses caractéristiques, puis je décrirai l'état des connaissances sur la formation du bois chez les Angiospermes et les acteurs moléculaires déjà identifiés. Cette partie s'achèvera par l'exposé des objectifs de thèse, de la stratégie adoptée et des méthodes expérimentales choisies. Un chapitre du manuscrit est dédié à la présentation du Matériel et des Méthodes. Les résultats de ma thèse sont ensuite structurés en 3 parties : choix des facteurs de transcription (FT) et propriétés, identification des cibles ADN des FT choisis par une approche *in vivo* génome-entier (ChIP-SEQ), et enfin, analyse de la fonction *in planta* des FT choisis par l'étude de peupliers transgéniques. Les résultats seront discutés au sein de chaque chapitre ; une conclusion et une discussion générale feront une synthèse des différents résultats.

Ce travail a été financé principalement par le projet Région Centre AMC2B, *Acteurs Moléculaires du Contrôle de l'Angle des Microfibrilles de Cellulose dans le Bois* et par le projet AIP Bioressources *REGEX*, ainsi que par le projet KBBE *TreeForJoules*. Ma bourse de thèse est une bourse mixte INRA EFPA / Région Centre.

Introduction et objectifs de thèse

Le bois est la plus importante source naturelle d'énergie renouvelable. En complément de son utilisation traditionnelle pour le chauffage, le bois constitue une nouvelle alternative aux combustibles fossiles. Le bois est en effet une source potentielle de bioéthanol, produit à partir de la cellulose contenue dans ses fibres. Il est utilisé également comme matériau (bois de construction, ameublement, emballages...) et comme matière première (bois de trituration pour la fabrication de papier, cartons, panneaux de particules...). Enfin, les cellules de bois en développement représentent un des plus importants réservoirs en CO₂ atmosphérique, réduisant ainsi l'un des acteurs majeurs du réchauffement climatique (Plomion *et al.* 2001, pour revue, voir Zhong *et al.* 2010a).

Mon travail de thèse s'inscrit dans une démarche plus large de décryptage des mécanismes moléculaires impliqués dans la formation du bois. Il porte sur l'étude fonctionnelle de 3 facteurs de transcription impliqués dans la formation du bois chez le peuplier, qui est à la fois un arbre modèle en biologie et une essence d'importance économique. Dans une première partie, j'exposerai les caractéristiques du peuplier, ses atouts biologiques et les ressources génétiques et génomiques disponibles. Puis, dans un second temps, je décrirai les différentes étapes de la formation du bois et des parois secondaires. Enfin, je me focaliserai sur les acteurs moléculaires déjà connus impliqués dans la formation du bois, avec une emphase particulière sur les facteurs de transcription.

1- Le peuplier

1.1- Le genre *Populus*

Taxonomiquement, le genre *Populus* appartient à la famille des Salicacées; il fait ainsi partie de la même famille que les saules et les osiers (genre *Salix*). Ce genre comprend 29 espèces de peuplier, définies principalement sur des critères morphologiques et écologiques, réparties en six sections (*Populus*, *Turanga*, *Aigeiros*, *Leucoides*, *Tacamahaca* et *Abaso*, Dillen *et al.* 2011). L'aire naturelle de répartition du genre *Populus* couvre la quasi-totalité de l'hémisphère Nord (Etats-Unis, Canada, Europe, Afrique du Nord, Russie, Chine et Japon). Aucun peuplier ne pousse à l'état naturel dans l'hémisphère Sud mais il est cultivé dans les régions tempérées de l'Argentine et du Chili (Dickmann & Kuzovkina, 2008). Les populations naturelles s'organisent le plus souvent le long des ripisylves et dans les zones

humides (Dickmann & Kuzovkina, 2008). Toutefois, certaines espèces telles que les trembles (*P.tremuloides*, *P. tremula*) se rencontrent en forêt. En France, les trois espèces autochtones sont *P. tremula* (tremble), *P. alba* (peuplier blanc) et *P. nigra* (peuplier noir).

Ces espèces sont en très grande majorité diploïdes ($2n = 38$), dioïques, avec des fleurs groupées en châtons. Elles arrivent à maturité sexuelle au bout de 4 à 8 ans, ce qui est précoce pour un arbre. Elles sont souvent caractérisées par une croissance rapide. Les peupliers sont peu longévifs, même si certains arbres remarquables peuvent atteindre 200 à 300 ans.

Certaines de ces espèces peuvent se croiser entre elles et avoir une descendance fertile, la barrière de l'espèce étant très faible dans le genre *Populus* (Stanton, 2009). L'hybridation interspécifique peut se produire naturellement entre espèces partageant la même aire de répartition, mais peut également être provoquée par croisements contrôlés. Cette propriété a été valorisée en création variétale pour générer des peupliers hybrides à forte croissance (vigueur hybride).

1.2- Une essence d'intérêt économique

Avec moins de 250 000 ha de plantations en France, soit moins de 2% des surfaces boisées (FAO, 2008), le peuplier est la seconde essence feuillue à être récoltée, en volume, après le chêne (35% des surfaces boisées). Ce résultat s'explique par sa très forte productivité. Actuellement, les peupleraies sont principalement menées en plantation monoclonale à faible densité, avec une rotation de 15 à 20 ans. Le bois ainsi produit est apte à être utilisé en déroulage et sciage pour la confection d'emballages légers et de panneaux de contre-plaqué, qui sont les deux utilisations les plus rémunératrices pour le bois de peuplier. Alternativement, il peut être utilisé en bois de trituration pour les usines de pâtes à papier. Le peuplier est également une essence pouvant être menée en taillis à courte rotation (5-7 ans) ou très courte rotation (2-3 ans) dans des plantations à haute densité (10 000 tiges/ha) à des fins de production de biomasse (plaquettes, granulés) (FCBA, 2009). Il est fort probable que ce type de plantations, dédiées bioénergie, se développe dans les années à venir.

En terme de ressources génétiques, la populiculture française repose actuellement principalement sur des plantations monoclonales de peupliers hybrides de clones

euraméricains (*Populus deltoides* x *P. nigra*, clones Robusta, Koster, I-214, Triplo, Blanc du Poitou, Flevo, ...) ou de clones interaméricains (*Populus deltoides* x *P. trichocarpa*, clones Beaupré, Boelare, Unal, Hunegem, Raspalje), issus principalement des programmes d'amélioration des pays voisins, Belgique, Italie et Pays-Bas. Les hybrides issus des croisements présentent généralement une forte vigueur hybride qui se traduit par une excellente productivité (Marron *et al.* 2007; Dillen *et al.* 2009). En France, la création du Groupement d'Intérêt Scientifique (GIS) Peuplier en 2001, regroupant le FCBA, l'IRSTEA et l'INRA, a permis de structurer le programme français d'amélioration génétique du peuplier. Les premiers cultivars français homologués sont actuellement en cours de diffusion (homologation publiée, arrêté du 08 avril 2013, paru au JO du 20 avril 2013). Les critères de sélection portent sur la productivité, la forme et le degré de branchaison des arbres, ainsi que la résistance aux pathogènes (rouille foliaire, brunissure des feuilles et chancre bactérien) (Berthelot *et al.* 2005).

Les peupleraies sont majoritairement localisées dans les plaines alluviales, le peuplier nécessitant beaucoup d'eau pour une productivité maximale. Les grandes régions populières sont la Picardie et le bassin Parisien, la région Poitou-Charentes et le bassin Aquitain, et la région Centre et la basse vallée de la Loire.

1.3- Un arbre modèle en biologie

En plus de leur intérêt économique, les peupliers possèdent des atouts biologiques qui ont conduit à leur adoption comme modèle d'étude, depuis plusieurs décennies, dans les laboratoires développant des recherches plus fondamentales, portant sur des processus biologiques, spécifiques ou non des arbres, tels que la formation du bois, le développement du système racinaire, la résistance à la sécheresse, l'entrée en dormance des bourgeons, la résistance à des pathogènes, etc... Ils sont disponibles sous forme de clones, sont multipliés aisément par bouturage, ont une croissance rapide : on peut donc conduire des expérimentations sur des individus génétiquement identiques, sur un temps relativement court pour un arbre. La technique de la transformation génétique, technique incontournable pour l'étude de la fonction de gènes *in planta*, a été disponible rapidement via *Agrobacterium*

tumefaciens, mais reste limité à un petit nombre de clones (Leplé *et al.* 1992, Busov *et al.* 2005).

De nombreuses cartes génétiques sont disponibles pour différents pédigrées de cartographie chez différentes espèces. Depuis 1994 et la publication de la première carte génétique construite sur une famille F2 *P. trichocarpa* x *P. deltoides* (Bradshaw *et al.* 1994), d'autres cartes génétiques ont été développées utilisant différentes espèces parentes (pour exemple, Jorge *et al.* 2005, Cervera *et al.* 2001, Gaudet *et al.* 2008).

Le génome du peuplier est relativement petit : sa taille physique est estimée à 500 Mb, soit 4 fois celui d'*Arabidopsis thaliana*; il est 50 fois plus petit que le génome du pin maritime. Il a été choisi pour être le premier génome d'arbre à être séquencé. Plus particulièrement, c'est Nisqually-1, un individu femelle de l'espèce *P. trichocarpa*, qui a été séquencé en whole genome shotgun avec une profondeur de 7,5X par le Joint Genome Institute (Tuskan *et al.* 2006). Cet individu a été récolté sur les rives du fleuve Nisqually à l'Ouest de l'État de Washington aux Etats-Unis. Depuis, la séquence, l'assemblage et l'annotation ont été améliorés et sont disponibles sur Phytozome (<http://www.phytozome.net/>). La version 3 est actuellement disponible. Le séquençage est maintenant à une profondeur de 10X et 423 Mb sont disponibles, dont 388 Mb répartis sur les 19 premiers scaffolds, correspondant aux 19 chromosomes du peuplier. Les 35 Mb restant se répartissent sur 1446 scaffolds de plus petite taille. Dans cette version de l'annotation, 41 335 loci correspondant à des modèles de gènes ont été identifiés, avec potentiellement 73 013 transcrits ayant diverses fonctions biologiques putatives. Dans ce qui suit, nous allons nous focaliser sur les mécanismes de formation du bois d'une façon générale, et des parois secondaires plus particulièrement.

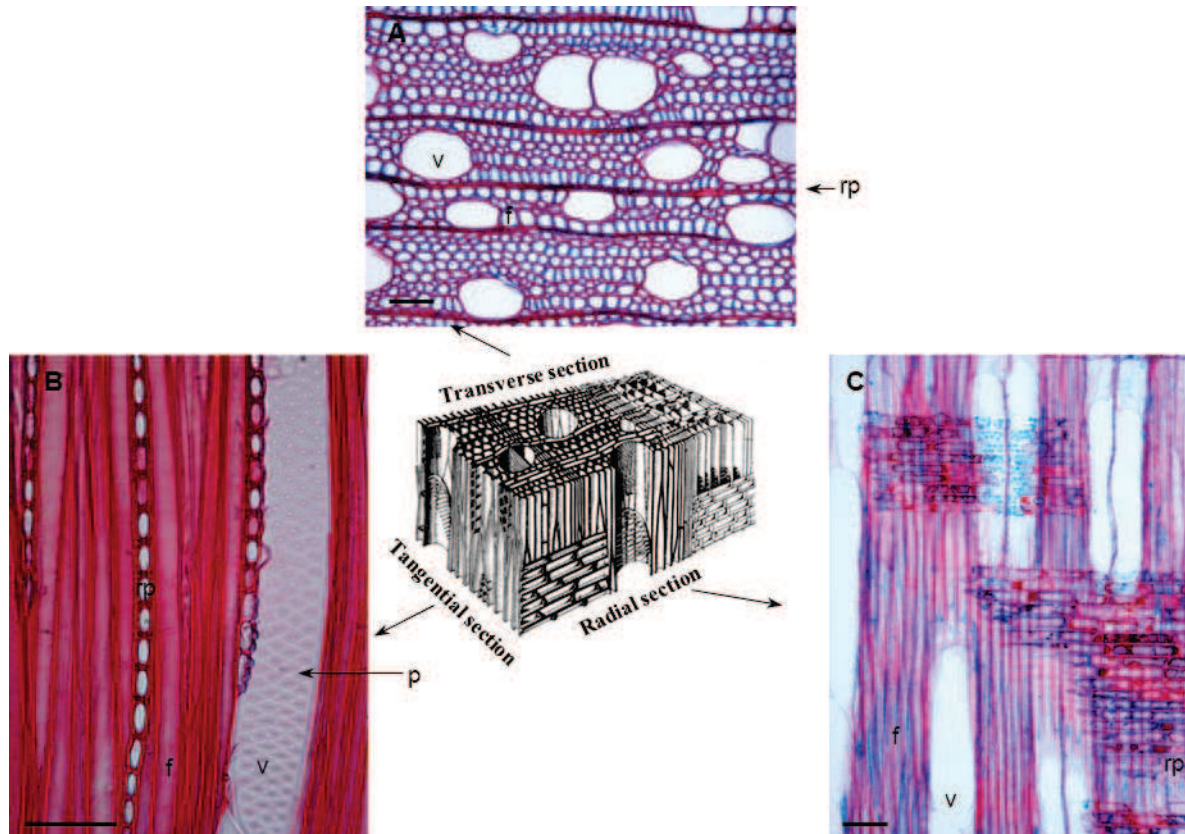


Figure 1 : Le bois résulte de l'association tri-dimensionnelle de différents types cellulaires.

Coupes transversales (A), tangentielle (B) et radiales (C) de bois de peuplier (*Populus tremula* x *P. alba*, clone INRA 717-1B4) colorées à la safranine et au bleu Astra. La safranine colore en rouge les parois lignifiées, alors que le bleu Astra colore en bleu les parois cellululosiques.

f: fibre; p: ponctuation; rp: rayons; v: vaisseau. La barre d'échelle correspond à 50 μm. (modifié à partir de Déjardin *et al.* 2010)

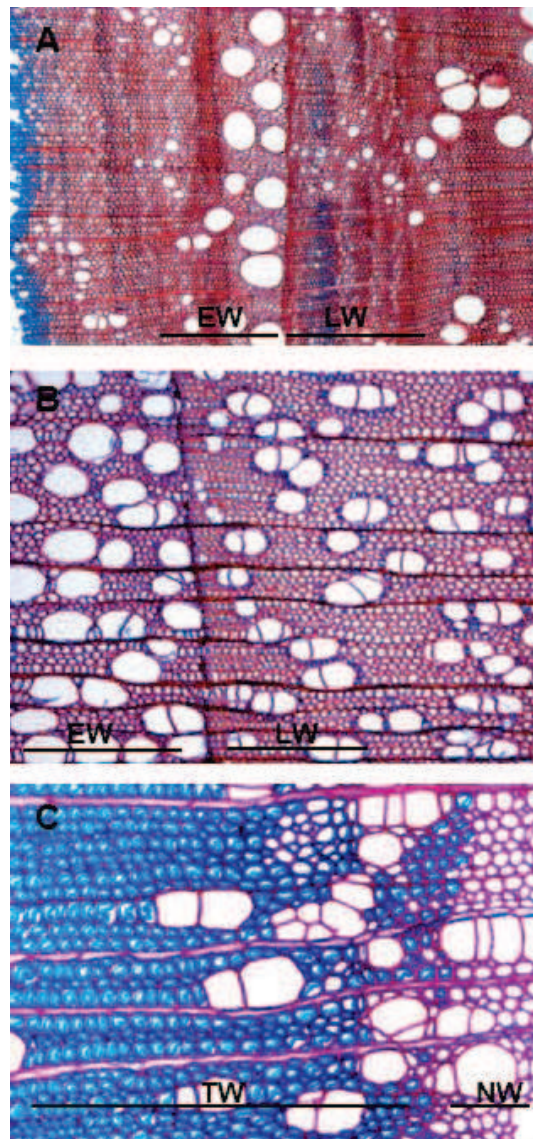


Figure 2 : Le bois est un tissu très variable.

Coupes transversales de bois de châtaignier (A) et de peuplier (B, C) colorées à la safranine/Bleu Astra, montrant des différences au niveau de la taille des vaisseaux, de leur nombre, de la taille des fibres, de la paroi secondaire des fibres entre bois de printemps (EW) et bois d'été (LW), bois normal (NW) et bois de tension (TW) (Déjardin *et al.* 2010).

2- Formation du bois et des parois secondaires

2.1-Anatomie du bois et fonctions pour l'arbre

Chez les plantes ligneuses, le bois ou xylème secondaire assure trois fonctions principales : (1) la conduction de la sève brute (l'eau et les solutés) des racines vers le reste de la plante, (2) le support et le maintien mécanique de l'arbre au cours de sa croissance et face aux différents stress environnementaux (vent, neige...) et (3) le stockage de réserves (Plomion *et al.* 2001, Demura & Fukuda. 2007, Déjardin *et al.* 2010). Chez les Angiospermes, différents types de cellules remplissent ces trois fonctions, on dit que le bois est hétéroxylé (fig.1). Les vaisseaux et les fibres sont impliqués, respectivement, dans la conduction de l'eau et le support mécanique, tandis que les cellules du parenchyme, organisées en rayons, sont impliquées dans le transfert des assimilats entre le phloème et le xylème, leur stockage temporaire sous forme d'amidon ou de lipides, et leur remobilisation à la nouvelle saison (Déjardin *et al.* 2010).

Le bois est un tissu très variable entre espèces, mais également au sein d'une même espèce ou chez un même individu (fig.2). Par exemple, au sein d'un cerne annuel, on distingue le bois de printemps (EW) produit lorsque les conditions de croissance sont favorables, en terme de température, disponibilité en eau et photopériode, du bois d'été (LW), produit plus tard dans la saison, plus dense, avec des fibres dont le lumen est réduit, avec une paroi très épaisse; les vaisseaux sont moins nombreux et de plus faible diamètre dans le bois d'été (fig.2A et 2B). Si on compare, à titre d'exemple, le bois de peuplier de celui de châtaignier, on observe des fibres dont le lumen est bien visible et des vaisseaux beaucoup plus nombreux (surtout au niveau du bois d'été) chez le peuplier (fig.2A et 2B) : son bois est moins dense que celui du châtaignier. De même, le bois de réaction est un bois particulier, produit cette fois-ci en réponse à des contraintes mécaniques (vent, neige, gravité), dont les propriétés mécaniques permettent à l'arbre de redresser ses axes, ou tout au moins de retrouver un certain équilibre. Chez les espèces à bois dur, le bois de réaction est appelé bois de tension (TW) car il se forme sur les faces supérieures des tiges inclinées où ses fibres sont sous fortes contraintes longitudinales (valeurs de $-3000 \mu\text{m/m}$, alors qu'on enregistre des valeurs de $-300 \mu\text{m/m}$ dans le bois normal de peuplier, Clair *et al.* 2006). Au niveau anatomique, il se caractérise par des fibres particulières, appelées fibres G, présentant une couche très épaisse d'aspect gélatineux (couche G), riche en cellulose (fig.2C) (Jourez *et al.*

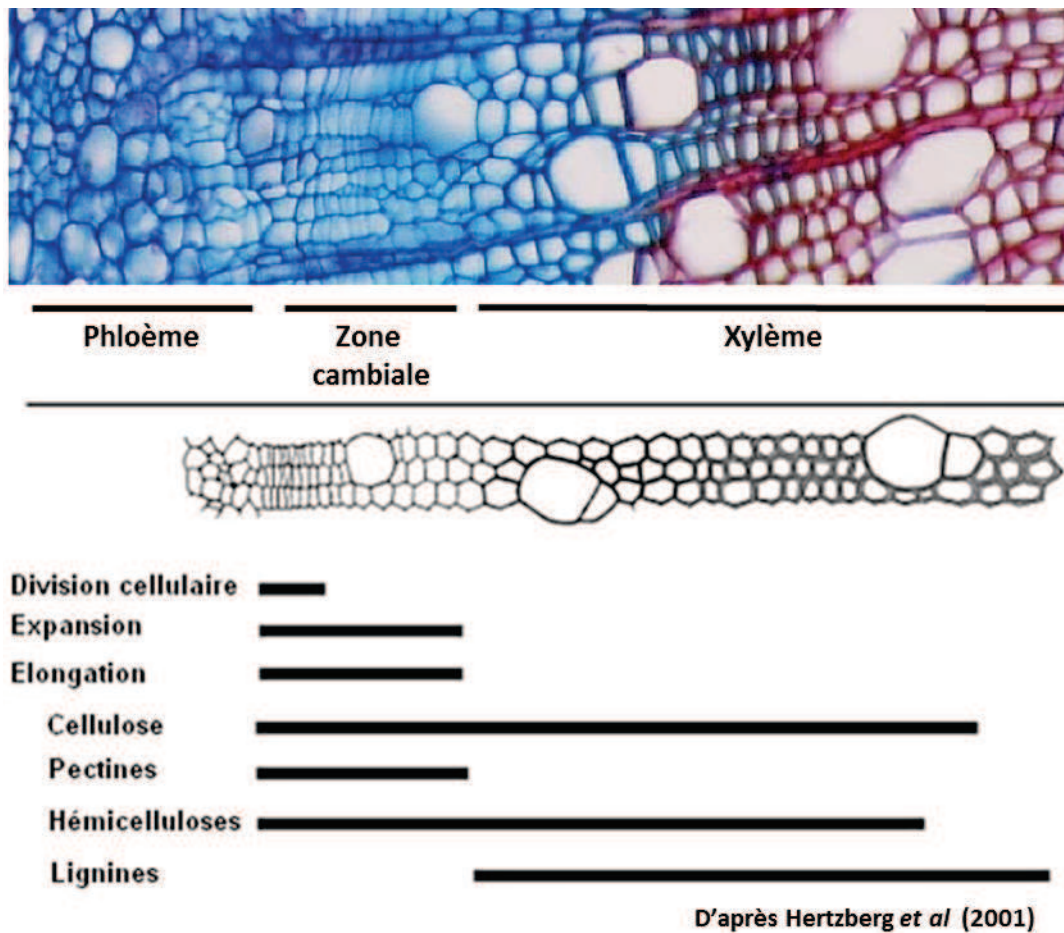


Figure 3 : Étapes de la formation du bois.

Les différentes étapes de la formation du bois (division, expansion et élongation cellulaire, synthèse des composés pariétaux) sont représentées spatialement sur une coupe transversale de peuplier - schéma ou coupe de tige de colorée à la Safranine/Bleu alcian. D'après Hertzberg *et al.* 2001.

2001). Dans les tiges inclinées artificiellement, les fibres G dans le TW sont formées de manière continue, alors qu'elles se forment plutôt de manière intermittente dans le TW formé naturellement (Tsai *et al.* 2012). Par rapport au bois opposé (OW), présent sur la face inférieure des tiges inclinées, le bois de tension présente moins de vaisseaux, de plus faible diamètre, mais plus de rayons ligneux (Jourez *et al.* 2001). Les vaisseaux et les fibres sont plus longs dans le bois de tension (Jourez *et al.* 2001, Aguayo *et al.* 2012).

2.2- Les différentes étapes de la formation du bois

Sous nos latitudes, le bois est produit de façon saisonnière en périphérie du tronc, grâce au fonctionnement d'un méristème secondaire, le cambium. La formation du bois implique différentes étapes principales (fig.3), comprenant d'abord des divisions cellulaires à partir du cambium. Dans le cas de la formation du bois de tension, on observe une stimulation de l'activité cambiale côté TW et une inhibition de son activité côté OW, conduisant à une excentricité de croissance (Mellerowicz & Sundberg 2008). Les initiales cambiales entament ensuite un processus d'expansion (élongation et élargissement radial) déterminant ainsi leurs dimensions finales avant le dépôt de la paroi cellulaire secondaire et sa lignification (Mellerowicz & Sundberg 2008). La lignification rend les parois rigides et imperméables, entraînant à terme la mort des cellules, fibres et vaisseaux. Le bois est donc formé principalement de l'enchevêtrement tridimensionnel de parois secondaires de cellules mortes, seules les cellules parenchymateuses des rayons restant vivantes. C'est pourquoi les propriétés du bois, mécaniques ou chimiques, sont très fortement liées aux propriétés des parois secondaires.

2.3. Propriétés et dépôt des parois secondaires

Les parois secondaires sont les principaux composants des éléments de vaisseaux et des fibres chez le bois des dicotylédones (Ye *et al.* 2006). Elles sont typiquement composées de cellulose, d'hémicelluloses (xylanes et glucomannanes) et de lignines (Zhong *et al.* 2007a).

La paroi cellulaire est composée de plusieurs couches qui sont déposées à différentes périodes pendant la différenciation cellulaire. La première couche se développant après la

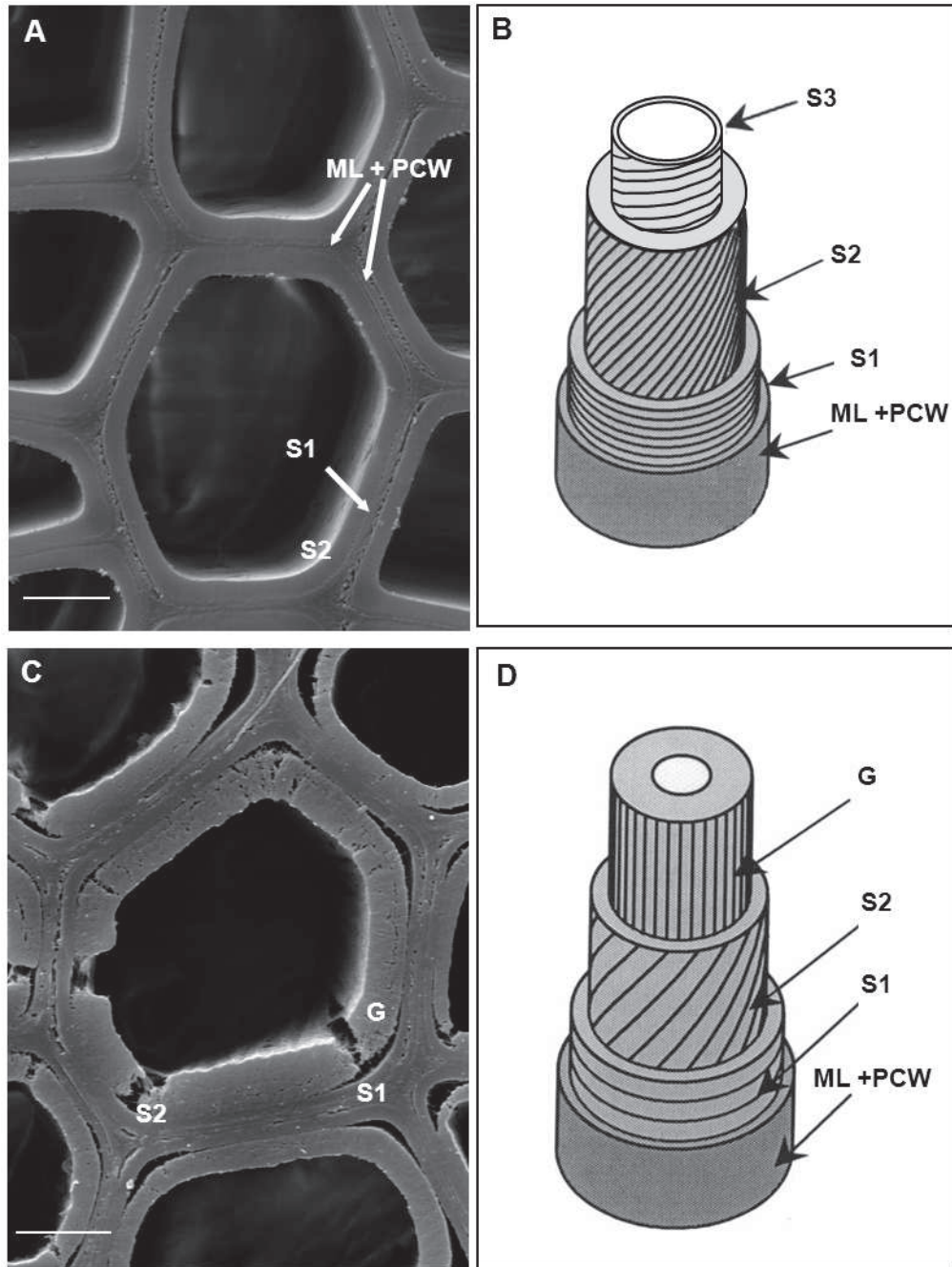


Figure 4 : La paroi secondaire est composée de plusieurs couches distinctes.

Microscopie à balayage électronique et schéma d'interprétation des différentes couches de la paroi cellulaire de fibres de bois normal (A, B) et de bois de tension (C, D) de peuplier. A noter que les lignes dans le schéma indiquent l'orientation préférentielle des microfibrilles de cellulose au niveau des différentes couches. ML: lamelle moyenne; PCW: paroi primaire; S1, S2, S3: couches de la paroi secondaire; G: couche gélatineuse dans les fibres de bois de tension. La barre d'échelle correspond à 5 μm (Déjardin *et al.* 2010).

division cellulaire est appelée lamelle moyenne ; elle se trouve entre les cellules de xylème et assure l'adhésion d'une cellule avec ses voisines. La lamelle moyenne a 0,5 à 1,5 μm d'épaisseur et est constituée de matières pectiques. Au début de la différenciation cellulaire, la paroi primaire se forme. Cette nouvelle couche fortement élastique est attachée à la lamelle moyenne et a 0,1 μm d'épaisseur. Elle est constituée de plusieurs couches de microfibrilles (MFs) de cellulose disposées de façon aléatoire au sein de cette paroi, au sein d'une matrice de substances pectiques et d'hémicelluloses. Quand la cellule en développement atteint sa taille définitive, la paroi secondaire est alors déposée, elle est beaucoup plus épaisse que la paroi primaire et joue un rôle important en termes de résistance mécanique (fig.4, Plomion *et al.* 2001). Elle est composée de 3 couches différentes, S1 (0,1 à 0,35 μm), S2 (1 à 10 μm) et S3 (0,5 à 1,10 μm). Chacune de ces couches est composée de MFs de cellulose arrangée d'une façon parallèle et dont l'angle avec l'axe de la fibre diffère d'une couche S à l'autre (fig.4). En particulier, l'alternance angle fort / angle faible permet d'augmenter la rigidité de la paroi et sa résistance mécanique. Du point de vue du biomécanicien, la paroi des cellules peut en effet être considérée comme un matériau composite, de type béton armé (Kerstens *et al.* 2001). Des modèles dérivés de la mécanique des matériaux composites ont mis en évidence l'importance de l'angle des microfibrilles de cellulose (AMF), et de son degré de cristallinité, pour expliquer les variations macroscopiques des propriétés mécaniques du bois (Fournier *et al.* 1994). Enfin, en dernière étape, la paroi des fibres et des vaisseaux est lignifiée. La lignification de la paroi commence à partir des coins de la cellule puis progresse de la lamelle moyenne vers la paroi primaire et la paroi secondaire (Donaldson *et al.* 2001). La lignification est également un paramètre important dans les propriétés physiques des parois car elle va les rigidifier et les imperméabiliser.

Les fibres de bois de tension présentent une couche G très particulière, qui remplace la couche S3 des fibres normales (Daniel *et al.* 2006, fig.4). Cette couche G est constituée principalement de cellulose très pure, organisée en macrofibrilles de 30 à 40 nm de diamètre (Daniel *et al.* 2006). Elle contient également quelques traces de monolignols (Joseleau *et al.* 2004; Gierlinger et Schwanninger, 2006; Lehringer *et al.* 2008) et d'autres polysaccharides, dont la nature est sujet à débat : alors que des xyloglucanes sont détectés dans les couches G de fibres matures de *P. tremula* x *P. tremuloides* (Nishikubo *et al.* 2007), on ne trouve pas de xyloglucanes dans les couches G de fibres en différenciation chez le clone INRA 717-1B4 (*P. tremula* x *P. alba*), mais une quantité importante de pectines de type rhamnogalacturonane I

(Guedes, 2013). Les microfibrilles de cellulose, beaucoup plus cristallines que dans les autres couches de la paroi secondaire, sont orientées dans le sens de l'axe de la fibre, avec un AMF très proche de 0° (Norberg et Meier, 1966).

3- Acteurs moléculaires connus de la formation du bois et des parois secondaires

Il est évident que pour l'étude de la formation du bois, les modèles ligneux sont préférables. Cependant, des avancées notables sur l'identification des mécanismes moléculaires impliqués dans la formation du bois ont été également obtenues à partir de travaux menés sur des plantes herbacées telles que la plante modèle *Arabidopsis thaliana* car elle peut développer sous certaines conditions de culture un cambium conduisant à la formation d'un xylème secondaire (Chaffey *et al.* 2002). Les connaissances actuellement acquises proviennent, d'une part, de travaux développant des approches de génomique, sans *a priori*, réalisés non seulement sur des arbres (Sterky *et al.* 1998, Hertzberg *et al.* 2001 ; Déjardin *et al.* 2004 ; Paux *et al.* 2005, Andersson-Gunnerås *et al.* 2006) mais aussi sur des plantes non ligneuses comme *Arabidopsis* (Ko *et al.* 2006) et le riz (Ambavaram *et al.* 2011) par exemple. D'autre part, ces connaissances viennent d'études avec une approche gène candidat que ce soit chez les arbres (Legay *et al.* 2010 ; Han *et al.* 2011 ; Zhong *et al.* 2011 ; Zhu *et al.* 2013) ou les plantes herbacées (Zhong *et al.* 2007a ; Fornale *et al.* 2010 ; Guillaumie *et al.* 2010 ; De Micco *et al.* 2012 ; Bhargava *et al.* 2013).

3.1- Enzymes et gènes de structure

Plusieurs travaux récents ont mis en évidence divers gènes impliqués dans la formation du bois. La formation de la paroi secondaire est pilotée par l'expression coordonnée de nombreux gènes spécifiquement impliqués dans la biosynthèse, l'assemblage et le dépôt de polysaccharides (cellulose, hémicelluloses, pectines), de lignines et de protéines dans la paroi cellulaire (Plomion *et al.* 2001, Zhong *et al.* 2007a, Du et Groover 2010).

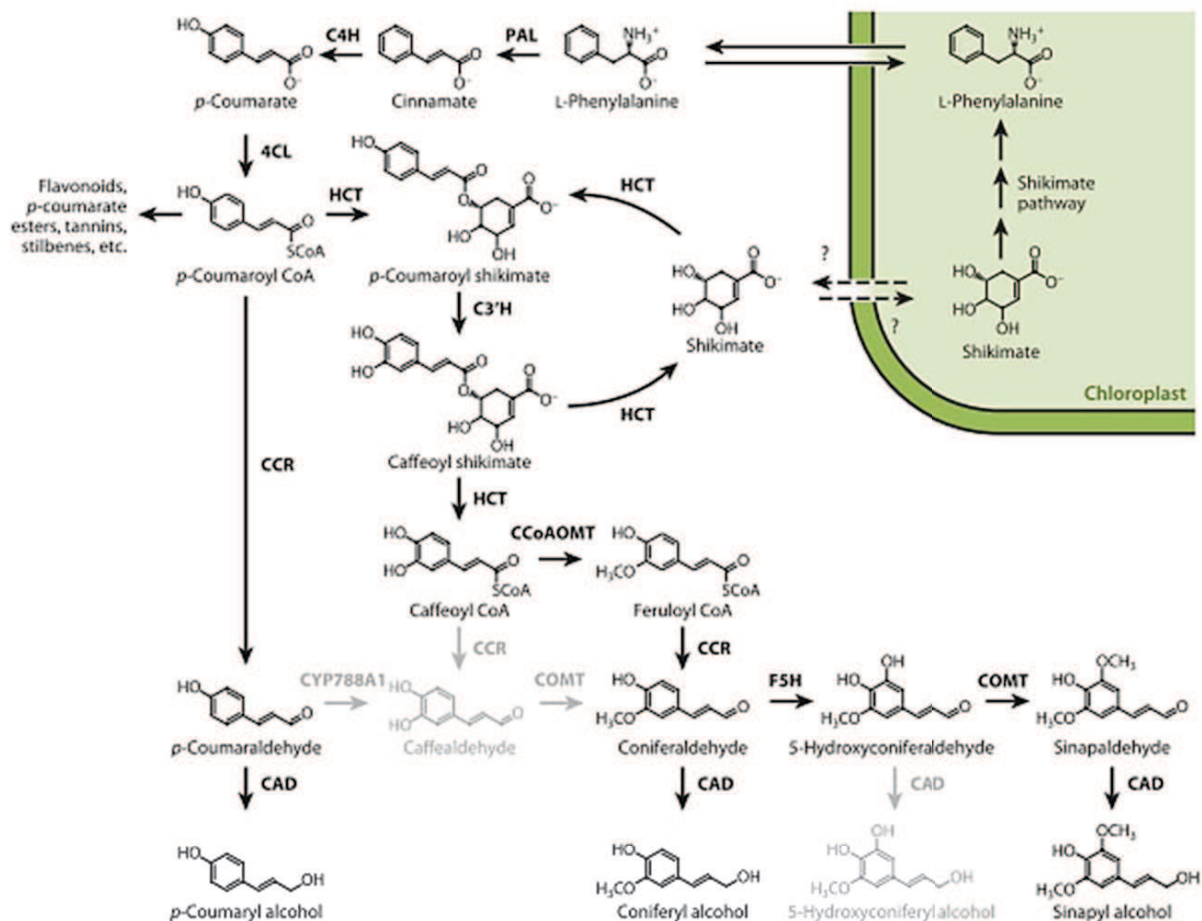


Figure 5 : Biosynthèse des monolignols *via* la voie des phenylpropanoïdes.

Sont indiqués les enzymes et les intermédiaires de biosynthèse des monolignols. Chaque flèche correspond à une réaction dans cette voie et à côté de chaque flèche se trouve le nom de l'enzyme qui catalyse la réaction en question. Les flèches et les intermédiaires gris clair indiquent des voies biochimiques qui ne sont probablement pas impliquées de façon significative dans la biosynthèse des lignines chez les plantes sauvages (Bonawitz & Chapple 2010).

3.1.1-Gènes de la voie de biosynthèse des lignines

Les lignines constituent l'un des polymères importants qui composent la paroi secondaire chez les plantes. Toutefois, elle est considérée comme un composé indésirable pour plusieurs utilisations notamment la production de biocarburants. Sa biosynthèse a été largement étudiée surtout chez les arbres. La lignine est synthétisée par couplage oxydatif de monomères d'alcools p-hydroxycinnamyliques, appelés monolignols. Ces derniers diffèrent suivant le nombre de méthylation au niveau de leur cycle aromatique. Après leur synthèse, ils sont transportés vers la paroi cellulaire où ils sont oxydés par des peroxydases et/ou des laccases pour être polymérisés. Les lignines des angiospermes contiennent principalement des unités guaiacyl (G) et syringyl (S) qui dérivent respectivement des alcools coniféryliques et sinapyliques (fig.5). Ces polymères se lient de façon covalente aux hémicelluloses donnant ainsi force et rigidité à la paroi cellulaire. Pendant les deux dernières décennies, la biosynthèse des lignines a été mieux comprise grâce à des études de mutants et/ou de plantes transgéniques (pour revue voir Boerjan *et al.* 2003; Vanholme *et al.* 2008).

Plusieurs enzymes interviennent dans la voie de biosynthèse des lignines telles que l'acide caféique O-méthyltransférase (COMT), l'alcool cinnamylique déshydrogénase (CAD), la cinnamoyl CoA réductase (CCR), la caffeoyl-CoA 3-O-méthyltransférase (CCoAOMT), la 4-coumarate-3-hydroxylase (C3H), la cinnamate-4-hydroxylase (C4H), la ferulate hydroxylase (F5H), la phenylalanine ammonia-lyase (PAL) et l'hydroxycinnamate-CoA-ligase (4CL) (pour revue, voir Bonawitz & Chapple, 2010, Boerjan *et al.* 2003). Par exemple, Leplé *et al.* (2007) ont montré l'implication de la CCR dans la biosynthèse des lignines au niveau de la paroi secondaire des cellules du xylème chez le peuplier. De surcroît, Prashant *et al.* (2011) ont montré que *LiCCR* est fortement impliqué dans la biosynthèse des lignines et possède un rôle important dans la détermination de la qualité et la quantité des lignines ainsi que dans la croissance et le développement des plantes. Par ailleurs, une analyse microarray chez le peuplier (*P. tremula* x *P. tremuloides*, clone T89) a révélé que les gènes codant pour les enzymes de la biosynthèse des lignines et de la cellulose, ainsi qu'un certain nombre de facteurs de transcription et d'autres régulateurs potentiels de la xylogénèse sont sous une régulation transcriptionnelle stricte qui dépend du stade de développement. Cette étude a permis ainsi de mettre en évidence et/ou de confirmer l'implication de plusieurs enzymes dans le processus de lignification, notamment une 4CL, deux CADs de type 2, deux

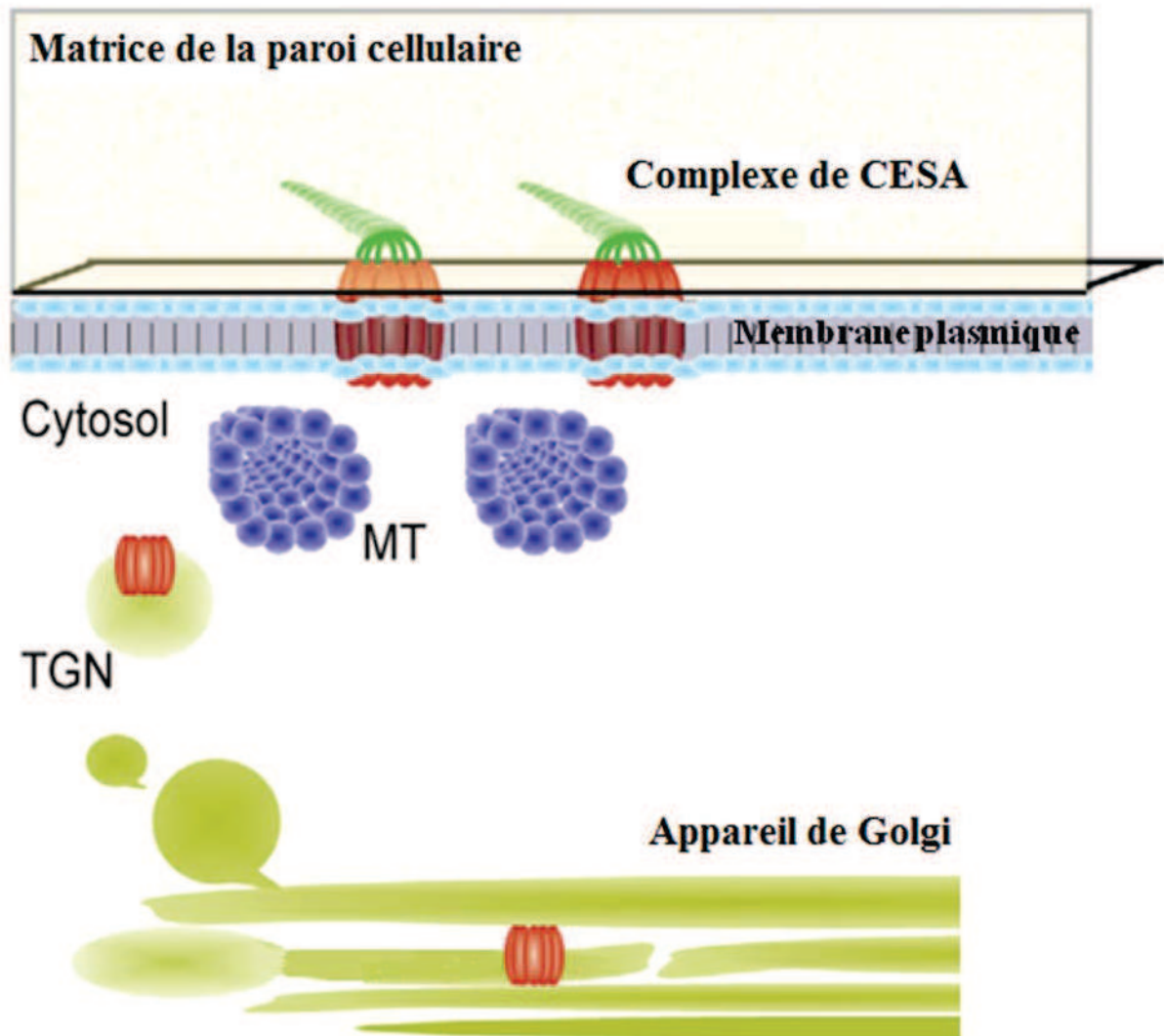


Figure 6 : Le complexe de cellulose synthase (CESA) dans un contexte cellulaire.

Le complexe de CESA (rouge) est transporté vers la membrane plasmique par exocytose à partir de l'appareil de Golgi (vert) à travers le réseau trans-Golgi (TGN). Durant la synthèse de cellulose, le complexe de CESA peut être guidé par les microtubules (bleu; MT). D'après Mutwil et al. 2008.

peroxydases et deux laccases, régulées positivement lors de la lignification (Hertzberg *et al.* 2001). Dans le bois de tension, dont les fibres ont une couche G pratiquement dépourvue de lignines, l'expression de plusieurs gènes de la voie de biosynthèse des lignines (principalement *CCoAOMT*, *PAL1*, *CCR*, *F5H*, *COMT*) est effectivement fortement réprimée dans le bois de tension de peuplier, par comparaison au bois normal (Andersson-Gunnerås *et al.* 2006).

3.1.2- Gènes de la voie de biosynthèse de cellulose et des hémicelluloses

En plus des lignines, la cellulose et les hémicelluloses sont des composants importants de la paroi secondaire.

Chez les plantes, la cellulose est synthétisée à partir de rosettes localisées au niveau de la membrane plasmique (fig.6). Chaque rosette est constituée de six globules, formés chacun par six unités catalytiques de cellulose synthase (CESA) (fig.6, Mutwil *et al.* 2008). Dix et dix-huit gènes CESA ont été identifiés respectivement chez *Arabidopsis* et *Populus trichocarpa*. Plusieurs études ont montré que 3 protéines CESA identifiées chez *Arabidopsis* (AtCESA4, 7 et 8) ainsi que leurs orthologues chez les arbres, sont impliquées dans la synthèse de cellulose, spécifiquement dans la paroi secondaire. Elles sont différentes des CESA impliquées dans la formation de la paroi primaire (pour revue voir Mellerowicz & Sundberg 2008).

Les glucuronoxylanes sont les hémicelluloses les plus importantes de la paroi secondaire des fibres et des éléments de vaisseaux chez les plantes dicotylédones. Plusieurs études ont permis d'identifier un grand nombre de gènes codant potentiellement pour des enzymes de biosynthèse des xylanes (pour revue voir Mellerowicz & Sundberg 2008). Certaines glycosyl transférases (GTs), spécifiquement exprimées dans le xylème de peuplier, semblent être essentielles pour la formation de la paroi secondaire au niveau de la synthèse des polysaccharides pendant la formation du bois (Brown *et al.* 2005). De plus, plusieurs familles de GTs impliquées dans la biosynthèse des xylanes ont été identifiées chez *Arabidopsis* (GT43, GT47, GT8) et le peuplier (GT43B, GT47C, GT8D et GT8E/D) (pour revue voir Mellerowicz & Sundberg 2008). Lee *et al.* (2010) ont montré que les 4 membres de

la famille GT43 d'*Arabidopsis*, I9H, IRX9, I14H et IRX14 sont impliqués dans la biosynthèse de glucuronoxylanes.

Enfin, plusieurs autres gènes codant pour des CAZYmes (carbohydrate active enzymes), impliqués dans la biosynthèse de la paroi cellulaire du bois ont été identifiés chez le peuplier et d'autres espèces ligneuses. Parmi ces gènes, on trouve ceux codant pour des CSLs (cellulose synthase-like), PME (pectin methyl esterases), XTH (xyloglucan endotransglucosylases XETs et hydrolases XEHs) et pour des expansines (pour revue, voir Mellerowicz & Sundberg 2008).

Un certain nombre de gènes codant pour des cellulose synthases, *PttCesA3-2*, *PttCesA8-2* (Andersson-Gunnerås *et al.* 2006), *EgCesA1* (Paux *et al.* 2005) et *EgCesA2* (Qiu *et al.* 2008) sont régulés positivement dans le bois de tension de peuplier ou d'eucalyptus. De même, plusieurs *XTH* (xyloglucan endotransglucosylase/hydrolase) sont régulés positivement pendant la formation de la couche G (Mellerowicz *et al.* 2008).

3.2- Facteurs de transcription impliqués dans la formation du bois et de la paroi secondaire

L'expression coordonnée des gènes impliqués dans la biosynthèse et l'assemblage des différents composants de la paroi nécessite une régulation au niveau transcriptionnel, passant au moins en partie par des FT, fonctionnant en cascade (Demura & Fukuda 2007, Zhong *et al.* 2007a). Les FT potentiellement impliqués appartiennent à différentes familles.

3.2.1- Facteurs de transcription à MADS-box

Plusieurs gènes codant pour des FT de plusieurs familles sont exprimés principalement pendant la formation du bois chez plusieurs espèces (Demura & Fukuda 2007). En effet, une étude des séquences EST de peuplier a montré que les gènes MADS-box (gènes possédant des domaines MADS-box qui sont des domaines de liaison à l'ADN) sont exprimés dans le bois et les tissus cambiaux, spécifiques aux plantes ligneuses (Leseberg *et al.* 2006).

3.2.2- Facteurs de transcription LBD

Des études récentes ont permis de mettre en évidence chez le peuplier un gène *PtaLBD1* (famille des facteurs de transcription LATERAL ORGAN BOUNDARIES DOMAIN (LBD)) de forte similarité avec le gène *LBD1* d'*Arabidopsis thaliana*. Ce gène codant pour un facteur de transcription est fortement exprimé dans le phloème et la zone cambiale. Il réprime l'expression de deux gènes clés *Class I KNOTTED1-like homeobox* qui favorisent l'identité du méristème dans le cambium et qui sont *ARBORKNOX1* (*ARK1*) et *ARBORKNOX2* (*ARK2*). Ce gène *PtaLBD1* ainsi que trois autres gènes *LBD* (*PtaLBD4*, *PtaLBD15* et *PtaLBD18*) sont principalement exprimés dans des tissus formant du bois suggérant un rôle de régulation de ces FT pendant la croissance secondaire dans le bois de peuplier (Yordanov *et al.* 2010). *PtrLBD15* est activé de façon directe par des PtrWNDs (wood-associated NAC domain transcription factor) (Zhong *et al.* 2011).

3.2.3- Facteurs de transcription de la famille MYB

Les FT MYB appartiennent à un groupe large de FT impliqués dans le développement des plantes et la différenciation de plusieurs types de tissus. Ils contiennent un motif de liaison à l'ADN appelé domaine MYB localisé dans le domaine N-terminal de la protéine. Les FT MYB sont classés en 3 sous-familles : MYB1R, MYBR2R3 et MYB3R en fonction du nombre de répétitions, dans le domaine de liaison à l'ADN, de 50 acides aminés. Chez les plantes, on trouve surtout des MYBs appartenant aux sous-familles MYB1R et MYBR2R3 (pour revue, voir Wilkins *et al.* 2009). Les FT MYB appartenant à la sous-famille R2R3 sont classés en 22 sous-groupes basés sur la présence de motifs distincts à l'extérieur des domaines MYB, qui sont, eux, très conservés (Kranz *et al.* 1998).

3.2.3.1- Certains MYB sont des régulateurs majeurs de la formation des parois secondaires

AtMYB46 et AtMYB83 fonctionnent de façon redondante dans la régulation de la biosynthèse de la paroi secondaire dans les fibres et les vaisseaux. Ce sont des régulateurs majeurs car ils activent une batterie de FT impliqués dans la régulation de plusieurs métabolismes clés (lignines, cellulose et xylanes). Il a été montré qu'ils activent les mêmes

FT en aval et sont essentiels pour un épaissement normal de la paroi secondaire (McCarthy *et al.* 2009). Zhong & Ye (2012) ont montré que MYB46 et MYB83 se fixent sur les promoteurs des gènes/FT cibles au niveau d'un site de 7 pb, appelé SMRE (SSecondary wall MYB-Responsive Element), dont la séquence est ACC(A/T)A(A/C)(T/C) et activent directement la transcription de FT (*AtMYB4*, *AtMYB7*, *AtMYB6*, *AtMYB32*, *AtMYB43*, *AtMYB52*, *AtMYB54*, *AtMYB58*, *AtMYB63*, *AtMYB85*, *AtKNAT7*, *AtC3H14*) mais aussi de gènes impliqués dans la biosynthèse des lignines (*CCoAOMT1*, *CAD*, *4CL3*, *Laccase12*, *Laccase11*, *Cinnamoyl-CoA reductase*, *ABC transporter*, *PAL1*, *PAL2* et *PAL4*) et des xylanes (*UXS3*, *I14H*, famille *GT8*, *GT14*, *GT4*). Ainsi, MYB46 et MYB83 jouent un rôle d'activateur dans la voie de biosynthèse des 3 principaux composants de la paroi secondaire : les lignines, les xylanes et la cellulose (McCarthy *et al.* 2009, Zhong & Ye 2012). Ils sont, à leur tour, régulés positivement par SND1, NST1, VND6, VND7 et PtrWNDs (Zhong *et al.* 2006 ; Zhong *et al.* 2008 ; Zhong *et al.* 2011).

Deux orthologues fonctionnels d'*AtMYB46* et *AtMYB83* chez le peuplier, *PtrMYB3* et *PtrMYB20*, régulent aussi la formation de la paroi secondaire et donc la voie de biosynthèse de cellulose, lignines et xylanes (McCarthy *et al.* 2010). *EgMYB2* est un orthologue fonctionnel d'*AtMYB46* chez *Eucalyptus*, capable de compléter le double mutant *myb46-myb83* chez *Arabidopsis* (pour revue, voir Zhong *et al.* 2010a). Il a d'abord été montré qu'*EgMYB2* était un activateur transcriptionnel de certains gènes de la voie de biosynthèse des lignines, notamment *EgCCR* et *EgCAD2* (Goicoechea *et al.* 2005). Dans des protoplastes d'*Arabidopsis*, *EgMYB2* est capable d'activer des gènes impliqués dans les 3 métabolismes (lignines, cellulose, xylanes), ce qui suggère très fortement qu'*EgMYB2* serait également un régulateur majeur de la formation de la paroi secondaire chez l'eucalyptus, à l'instar d'*AtMYB46* (Grima-Pettenati *et al.* 2012).

3.2.3.2- Facteurs de transcription MYB régulateurs positifs de gènes/FT de la voie de biosynthèse des lignines

Plusieurs MYBs sont des activateurs directs de la biosynthèse des lignines (Zhong *et al.* 2009). En effet, plusieurs régulateurs transcriptionnels de cette famille, comprenant *AtMYB61*, *AtMYB68* et *AtMYB63*, activent la biosynthèse de ce polymère (Zhou *et al.*

2009 ; Wang *et al.* 2010). AtMYB61 est impliqué dans le contrôle de la lignification chez *Arabidopsis*. Il est, de plus, capable de réguler directement l'expression d'*AtKNAT7* et *AtCCoAOMT* (Romano *et al.* 2012). Par ailleurs, MYB103 est nécessaire à l'expression de *F5H* et à la biosynthèse des lignines de type S (syringyl) au niveau de la hampe florale d'*Arabidopsis* (Öhman *et al.* 2013). De même, NtMYBJS1 chez le tabac (Gális *et al.* 2006), VvMYB5a chez la vigne (*Vitis vinifera*) (Deluc *et al.* 2006), PtMYB1 et PtMYB4 chez le pin (*Pinus taeda*) (pour revue, voir Zhong *et al.* 2010a) et PtrMYB28 chez le peuplier (Karpinska *et al.* 2004 ; pour revue, voir Zhong *et al.* 2010a) régulent positivement la biosynthèse des lignines.

De plus, deux FT homologues chez *Arabidopsis*, MYB58 et MYB63, sont exprimés dans les fibres et les vaisseaux. Ils sont sous le contrôle de SND1, MYB83 et MYB46. MYB58 et MYB63 peuvent réguler directement et spécifiquement l'expression des gènes de la biosynthèse des lignines ainsi qu'un gène (*LAC4*), codant pour une laccase associée à la paroi secondaire; seul MYB58 a la possibilité de se lier directement sur son promoteur et l'activer (Zhou *et al.* 2009). Chez le peuplier, un FT PtrMYB28, exprimé dans le bois au cours de la formation de la paroi secondaire, induit l'expression d'un gène (*4CLI*) impliqué dans la biosynthèse des lignines pendant la formation du bois (Zhong *et al.* 2009).

3.2.3.3- Facteurs de transcription MYB régulateurs négatifs de gènes/FT de la voie de biosynthèse des lignines

Inversement, plusieurs FT de type MYB, appartenant principalement au sous-groupe 4 présentant un domaine de répression EAR (Ethylene-responsive element binding factor associated amphiphilic repression) (Kazan 2006), agissent comme des régulateurs négatifs de la biosynthèse des lignines. AmMYB308 et AmMYB330 sont des répresseurs transcriptionnels de gènes de biosynthèse des monolignols chez le mufler (*Antirrhinum majus*) (Tamagnone *et al.* 1998). Chez le blé (*Triticum aestivum*), TaMYB4, un FT spécifique des monocotylédones, pourrait réguler négativement l'expression des gènes *CAD* et *CCR* (Ma *et al.* 2011). Il en est de même pour ZmMYB31 et ZmMYB42 chez le maïs (*Zea mays*) (Fornale *et al.* 2010; Sonbol *et al.* 2009) et PvMYB4 chez le panic érigé (*Panicum virgatum*) (Shen *et al.* 2012). De même, bien que la *CCoAOMT* soit régulée positivement dans des

plants d'*Arabidopsis* sur-exprimant *AtMYB4* (Jin *et al.* 2000) et des plants de tabac surexprimant *LlMYB1* du faux mimosa (*Leucaena leucocephala*) (Omer *et al.* 2013), d'autres gènes de biosynthèse des lignines (*C4H*, *4CL3*, *4CL1*, *CAD*, *PAL* et *CHS*) sont régulés négativement chez ces plants, suggérant que *AtMYB4* et son orthologue *LlMYB1* pourraient être des répresseurs de la biosynthèse de ce polymère dans certains cas (Jin *et al.* 2000, Omer *et al.* 2013). Preston *et al.* (2004) ont identifié *AtMYB32* comme répresseur de la lignification chez *Arabidopsis*. EgMYB1 agit aussi comme régulateur négatif de la lignification ainsi que de la formation de la paroi secondaire chez le peuplier et chez *Arabidopsis* (Legay *et al.* 2007, Legay *et al.* 2010).

Un autre gène MYB *PttMYB21a* chez le peuplier est exprimé pendant la formation du bois, principalement lors de la formation de la paroi secondaire des fibres de xylème et du phloème, et la diminution de son expression par stratégie antisens a conduit à la sur-expression d'un gène codant pour une CCoAOMT, une enzyme qui catalyse une étape de méthylation clé dans la biosynthèse des monolignols (Karpinska *et al.* 2004). *PttMYB21a* est régulé positivement dans le bois de tension. Il est proposé que ce FT soit un répresseur de la lignification.

3.2.4- Facteurs de transcription de type NAC

Les facteurs de transcription avec domaine NAC (*NAM*, *ATAF1/2* et *CUC2*) sont des régulateurs transcriptionnels spécifiques aux plantes et sont importants dans la croissance et le développement de celles-ci (pour revue, voir Zhong *et al.* 2010a). Plusieurs gènes codant pour des FT de cette famille sont exprimés préférentiellement dans le bois en développement, notamment dans les vaisseaux en cours de différenciation. L'étude fonctionnelle de certains de ces gènes a mis en évidence de nouveaux régulateurs-clés de la formation du bois (Mitsuda *et al.* 2005). Sept gènes *VND* (*VASCULAR-RELATED NAC-DOMAIN*) sont exprimés dans des tissus vasculaires en développement, notamment *VND6* et *VND7* qui ont montré une expression spécifique dans les vaisseaux immatures où ils sont des facteurs-clés impliqués dans l'épaississement de la paroi secondaire (Mitsuda *et al.* 2005). Deux gènes appartenant à une sous-famille très proche de la sous-famille de *VND*, *NST1* et *NST2* (*NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1* et 2), induisent la formation de parois

secondaires épaisses lorsqu'ils sont exprimés de façon ectopique (Mitsuda *et al.* 2005). Ces deux gènes sont régulés par MYB26 dans les anthères d'*Arabidopsis* (pour revue, voir Zhong *et al.* 2007a). La surexpression d'un autre membre de cette famille, *NST3* ou *SND1*, induit également un épaississement de la paroi secondaire dans différentes cellules (Zhong *et al.* 2006, Mitsuda *et al.* 2007). Ainsi, ces gènes *NST* ont des fonctions-clés dans la régulation transcriptionnelle qui dirige la différenciation des fibres (Demura & Fukuda 2007) et/ou la biosynthèse de la paroi secondaire au niveau des fibres (*SND1*, *NST2* et *NST1*) (Mitsuda *et al.* 2005, pour revue, voir Zhong *et al.* 2010a). *NST1*, *NST2*, *SND1*, *VND6* et *VND7* fonctionnent au sommet du réseau de régulation de la formation de la paroi secondaire pour activer l'ensemble du programme de la biosynthèse de la paroi secondaire (Mitsuda *et al.* 2005, Mitsuda *et al.* 2007, Zhong *et al.* 2006, Zhong *et al.* 2007b, Zhong *et al.* 2010a).

SND1 régule positivement *PtrMYB021* (Li *et al.* 2012), ainsi que plusieurs autres FT (Zhong *et al.* 2006, Zhong *et al.* 2008) dont *MYB46* et *MYB83* qui sont des cibles directes de *SND1* mais aussi de *NST1*, *VND6* et *VND7* (Zhong *et al.* 2007c, McCarthy *et al.* 2009). Un variant d'épissage alternatif de *PtrSND1* nommé *PtrSND1-A2^{IR}* code pour une protéine sans pouvoir de fixation et transactivation mais capable de se dimériser avec *PtrSND1* dans le noyau engendrant ainsi son piègeage; cela crée un dominant négatif qui module le réseau transcriptionnel de *SND1* (Li *et al.* 2012). *PtrMYB3* et *PtrMYB20* ainsi que d'autres FT chez *Arabidopsis* comme *SND3*, *MYB103*, et *KNAT7* (Knotted1-like homeodomain protein) (Zhong *et al.* 2007c, McCarthy *et al.* 2009) sont aussi des cibles directes de régulateurs clés de la paroi secondaire de type NAC. D'autres FT spécifiquement exprimés dans des cellules en train de déposer leur paroi secondaire chez *Arabidopsis* (*SND2*, *MYB20*, *MYB42*, *MYB52*, *MYB54*, *MYB58*, *MYB63*, *MYB69* et *MYB85*) sont régulés par *MYB46* et *MYB83* (McCarthy *et al.* 2009). *SND2* régule positivement l'expression de *MYB103* et *ANAC19* mais aussi celle de *SND1* bien que ce dernier le régule positivement à son tour. *SND2* active également *CesA4* et *FLA12* codant respectivement pour une cellulose synthase et une protéine à arabinogalactanes à domaine fascicline ainsi que d'autres gènes impliqués dans la polymérisation des lignines (Hussey *et al.* 2011). Chez le riz, *CCR*, 3 gènes *OMT* (O-méthyltransférase) et plusieurs gènes *WAK* (wall-associated kinases) sont régulés positivement suite à la surexpression d'*OsNAC9*, suggérant un rôle potentiel de ce dernier dans la formation de la paroi secondaire (Redillas *et al.* 2012). Chez une autre graminée, le panic érigé (*Panicum virgatum*), la surexpression d'un FT NAC d'*Arabidopsis*, *LONG*

VEGETATIVE PHASE ONE (AtLOV1) modifie la morphologie de la plante et engendre une légère mais significative augmentation du taux de lignines totales mais aussi une modification de la composition en monolignols des parois cellulaires (Xu *et al.* 2012).

Un groupe de FTs NAC a été identifié chez le peuplier et noté PtrWNDs, comprenant notamment PtrWND2B et PtrWND6B qui sont des orthologues fonctionnels de NST1 et SND1. Ils sont principalement exprimés dans le xylème de peuplier. Ces deux régulateurs transcriptionnels sont capables d'activer l'expression de gènes de biosynthèse de la paroi secondaire dont *CesA4*, *CesA7*, *CesA8*, *IRX8*, *IRX9*, *FRA8*, *4CL1* et *CCoAOMT1*. Ils fonctionnent aussi comme régulateurs de l'activation de la synthèse de la paroi secondaire (Zhong *et al.* 2010b). De plus, Zhong *et al.* (2011) ont montré que PtrWND2B, PtrWND6B et EgWND1 peuvent activer directement un grand nombre de FT impliqués dans la biosynthèse et la modification de la paroi secondaire (*PtrNAC118*, *PtrNAC150*, *PtrNAC156*, *PtrNAC157*, *PtrMYB3*, *PtrMYB18*, *PtrMYB21*, *PtrMYB74*, *PtrMYB75*, *PtrMYB121*, *PtrMYB128*, *MYB46*, *MYB83*, *MYB103*, *PtrKNAT7*, *KNAT7*, *PtrWRKY12*, *PtrWRKY13*, *SND3*, *PtrLBD15*, *PtrZF1* et *PtrGATA8*).

3.2.5- Facteurs de transcription de type KNOTTED-like (KNOX)

Li *et al.* (2012) a montré que KNAT7 (Knotted1-like homeodomain protein) régule négativement la formation de la paroi secondaire chez *Arabidopsis* mais aussi chez le peuplier en réprimant l'expression des gènes de biosynthèse de lignines (*PAL1*, *C4H*, *4CL-I*, *CCR1*, *HCT*, *C3H1*, *CCoAOMT1* et *COMT1*), de cellulose (*CesA7* et *CesA8*) et de xylanes (*IRX8*, *IRX9* et *FRA8*). Le gène codant pour ce FT ainsi que son homologue chez le peuplier *PtrKNAT7* sont des cibles directes de PtrWNDs, MYB46 et MYB83, et sont donc régulés positivement par ces protéines (Zhong *et al.* 2011 ; Zhong et Ye 2012). MYB75 et KNAT7 interagissent au sein d'un complexe multi-protéique et interviennent dans la régulation de la formation de la paroi secondaire au niveau de la hampe florale d'*Arabidopsis* (Bhargava *et al.* 2013).

BREVIPEDICELLUS (BP), l'un des 7 gènes *KNOX* chez *Arabidopsis*, est capable de se fixer sur les promoteurs de certains gènes de biosynthèse des lignines (Mele *et al.* 2003). Par ailleurs, chez le pêcher (*Prunus persica* L. Batsch), le gène *KNOPE1* appartenant à la

famille KNOX, est un répresseur de la lignification des cellules en système hétérologue en réprimant les gènes de biosynthèse des lignines comme *AtC4H*, *AtCL1*, *AtCAD4*, *AtLAC4* et *AtCCoAOMT* (Testone *et al.* 2012).

3.2.6- Facteurs de transcription de type WRKY

Les facteurs de transcription WRKY peuvent avoir également un rôle dans la formation des parois cellulaires. En effet, une mutation touchant ces FT chez *Arabidopsis* et *Medicago* a engendré une augmentation importante de la biomasse de la tige. De surcroît, *AtWRKY12* réprime l'expression de trois FT : *NST2* en se fixant directement sur son promoteur, *C3H14* et *C3H14L* (facteurs de transcription de type CCCH à doigt de zinc) sans liaison directe avec leurs promoteurs (Wang *et al.* 2010). *PtrWRKY12* et *PtrWRKY13* sont régulés positivement par *PtrWND*s et *EgWND1* (Zhong *et al.* 2011). Par ailleurs, Guillaumie *et al.* 2010 ont montré l'implication de *VvWRKY2* dans la biosynthèse des lignines et le développement du xylème. Les transcrits du gène *VvWRKY2* sont accumulés spécifiquement dans les tissus lignifiés (xylème, sclérenchyme). Ce FT peut se lier *in vitro* au promoteur de *VvC4H* et activer son expression. De plus, plusieurs gènes de biosynthèse de lignines : *C4H*, *4CL*, *HCT*, *C3'H*, *CCoAOMT*, *CCR*, *CAD* sont régulés positivement chez des plants de tabac surexprimant ce FT.

Par ailleurs, une étude du transcriptome a permis d'identifier 2 gènes *WRKY* : *WRKY20/PU07328* et *PU09430* qui sont, respectivement, régulés positivement et négativement durant la formation du bois de tension chez le peuplier (Andersson-Gunnerås *et al.* 2006).

3.2.7- Facteurs de transcription de type HD-ZIP classe III

L'expression du gène *PtaHB1* d'un *Populus tremula x Populus alba* (*Pta*) qui code pour une protéine HD-ZIP III est étroitement associée à la formation du bois (Ko *et al.* 2006). Par ailleurs, le gène *PtrHB7* de *Populus trichocarpa*, exprimé préférentiellement dans le cambium vasculaire, joue un rôle critique dans le contrôle de la balance de différenciation entre les tissus du phloème et du xylème secondaire d'une manière dose-dépendante. D'une part, il régule positivement non seulement l'expression de plusieurs FT impliqués dans la

formation de la paroi secondaire et la différenciation du xylème tels que *SND1*, *SND3*, *NST1*, *NST2*, *VND7*, *KNAT7*, *MYB20*, *MYB43*, *MYB46*, *MYB52*, *MYB85*, *MYB103* et *WND1B*, mais aussi celle de certains gènes de biosynthèse des lignines (*PAL1*, *4CL1*, *HCT*, *CAD5*), de cellulose (*CesA4*, *CesA7*, *CesA8*, *CesA10*) et d'hémicelluloses (*GT43B*, *IRX8*, *IRX9*, *IRX10*, *IRX14*, *IRX15*). D'autre part, c'est un régulateur négatif possible de *LBD1* et *ARBORKNOX2* (*ARK2*) (Zhu *et al.* 2013).

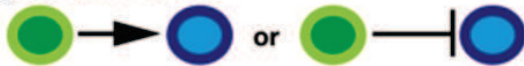
3.2.8- Facteur de transcription SHINE

Récemment, la surexpression d'un FT SHINE d'*Arabidopsis* (AtSHN) chez le riz, a entraîné une réduction importante de la quantité de lignines et, en même temps, une augmentation importante de la quantité de cellulose. Ce régulateur de transcription est capable de se fixer directement sur les promoteurs de plusieurs gènes MYB (*MYB58/63*, *MYB20/43*) et NAC (*NST1/2/SND1*, *VND4/5/6*). Sa surexpression a aussi modifié l'expression de nombreux gènes de biosynthèse de lignines (*CAD* et *4CL*), de cellulose et d'hémicelluloses (gènes *CESA* et *CESA-like*). Ainsi, ce FT est capable de réguler d'une manière coordonnée les voies de biosynthèse des lignines et de cellulose en réprimant certains activateurs-clés (NACs) et des MYBs spécifiques de la biosynthèse des lignines, tout en activant les MYBs et autres FT spécifiques de la biosynthèse de cellulose (Ambavaram *et al.* 2011).

4- Objectifs de la thèse

La formation du bois est un processus physiologique complexe, qui nécessite l'expression coordonnée dans l'espace et dans le temps d'un nombre important de gènes. Des approches globales d'analyse du transcriptome ont clairement mis en évidence cette complexité (Hertzberg *et al.* 2001, Déjardin *et al.* 2004, Andersson-Gunnerås *et al.* 2006). Cette expression coordonnée est sous le contrôle d'un nombre plus réduit de régulateurs, dont font partie une catégorie de gènes, codant pour des facteurs de transcription (FT). Les FT sont des protéines qui, en se fixant directement sur l'ADN au niveau de séquences régulatrices dans les régions promotrices des gènes, vont activer ou réprimer la transcription du/des gènes. L'interaction entre les protéines et l'ADN joue un rôle-clé dans la régulation

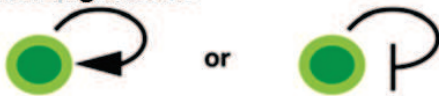
Régulation directe



Régulation indirecte



Autorégulation



Rétrocontrôle

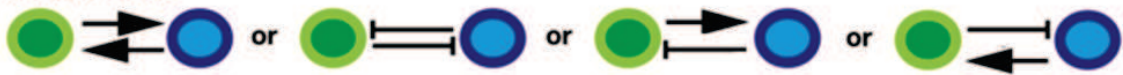


Figure 7 : Régulation transcriptionnelle par les FT.

Les flèches indiquent les différentes possibilités de régulation de la transcription des cibles directes (bleu) ou indirectes (rouge) par les FT (vert) ; les extrémités pointues correspondent à une régulation positive et les extrémités plates à une régulation négative. Schéma modifié à partir de Yant (2012).

des fonctions cellulaires telles que la transcription de gènes (Das *et al.* 2004). Les FT sont au cœur des réseaux de régulation transcriptionnelle, connectés à plusieurs autres FT et participant à des mécanismes de rétrocontrôle régulationnel (Yant 2012). Ils peuvent agir comme des activateurs ou des répresseurs de la transcription (ou les deux à la fois) d'une façon directe ou indirecte (Prouse et Campbell 2012) (fig.7). Ils peuvent parfois s'autoréguler positivement ou négativement, et un rétrocontrôle positif ou négatif peut avoir lieu entre ces protéines et leurs cibles (notamment d'autres FT).

Si la fonction de certains FT est maintenant bien connue, en particulier chez la plante-modèle *Arabidopsis*, comme décrit en partie introductive, il reste encore de nombreux points à élucider. Notamment, on ne connaît pas toujours les cibles directes des FT. Chez le peuplier, le nombre de FT déduit de la séquence de *P. trichocarpa* v3.0 est potentiellement de 4244, correspondant à 2455 modèles de gène (Plant Transcription Factor database v3.0, <http://planttfdb.cbi.pku.edu.cn/index.php?sp=Pth>), soit environ 6% des modèles de gènes ou des protéines prédites. Ce nombre élevé de FT souligne la grande plasticité du transcriptome de la cellule, mais également la redondance fonctionnelle possible entre FT.

Ma thèse vise à identifier différents acteurs majeurs de ces réseaux de régulation, FT et leurs cibles ADN, avec pour objectif principal, de mieux comprendre au niveau mécanistique la formation du bois. Pour répondre à cette question de recherche, j'ai développé une stratégie articulée autour de trois points majeurs, qui seront repris ultérieurement sous forme de trois chapitres de résultats. Pour un nombre limité de FT choisis, nombre compatible avec un travail de thèse, j'ai cherché à identifier leurs cibles ADN en utilisant une technologie nouvelle dans le domaine végétal, et permettant d'accéder aux cibles de FT, sans *a priori* et à l'échelle du génome (technologie ChIP-SEQ, immunoprécipitation de chromatine associée à un séquençage haut-débit des fragments ADN précipités). En parallèle, j'ai produit des peupliers transgéniques affectés pour l'expression de ces FT pour obtenir également d'autres informations sur la fonction de ces FT dans la formation du bois chez le peuplier. Je vais détailler la stratégie employée et les principaux résultats marquants.

4.1- Choix des facteurs de transcription

Le choix des FT est un élément important de la thèse. Le choix était relativement contraint par le fait que ce nombre devait être obligatoirement faible et que nous souhaitions y inclure un FT contrôle pour nos expériences, c'est-à-dire un FT impliqué dans la formation du xylème secondaire, et pour lequel au moins une cible avait été démontrée expérimentalement. Nous avons avant tout choisi des candidats expressionnels, c'est-à-dire des FT spécifiquement exprimés dans le xylème secondaire, et surexprimés dans le bois de tension. Nous avons en effet retenu le bois de tension comme modèle biologique d'étude car il possède des propriétés intéressantes pour différents aspects : c'est un bon modèle pour comprendre les mécanismes impliqués dans l'orientation des microfibrilles de cellulose - dont l'angle est proche de 0° dans les fibres G - ou le ratio lignines/cellulose, les fibres G étant pratiquement dépourvues de lignines. Au niveau expérimental, il est de plus facilement induit par inclinaison de tiges de peuplier : on peut ainsi comparer au sein de la même tige, deux types de bois, bois de tension et bois opposé, présentant des caractéristiques anatomiques et biochimiques très différentes.

Nous avons retenu 3 FT appartenant à la famille MYB : MYB090, proche d'AtMYB52 et AtMYB54, fortement exprimé dans le xylème secondaire et surexprimé dans le bois de tension ; MYB156 et MYB221, orthologues d'EgMYB1, répresseur de la transcription de certains gènes de la lignification, ayant pour cibles démontrées les promoteurs des gènes *CCR* et *CAD* (Legay *et al.* 2007). Les séquences ont été clonées et analysées, les FT sont correctement localisés dans le noyau ; des anticorps polyclonaux anti-MYB090 et anti-MYB221 ont été produits, ce sont des outils indispensables à la mise en œuvre de l'immunoprécipitation de chromatine (ChIP).

4.2- Identification des cibles ADN des FT à l'échelle du génome

Pour identifier les cibles ADN des FT, nous avons choisi de développer sur notre système d'étude la technique d'immunoprécipitation de chromatine, technique qui permet d'avoir accès aux interactions protéines/ADN *in vivo* (Zinzen *et al.* 2009). Associée à un séquençage haut-débit de séquences courtes de type Illumina, cette technique, appelée alors ChIP-SEQ, permet d'identifier les sites de fixation du FT au niveau génome entier, et avec une résolution dépendante de la profondeur de séquençage, pouvant atteindre 10 à 50 pb

(Valouev *et al.* 2008). Cette technique est particulièrement intéressante dans la mesure où elle permet d'identifier des cibles, sans *a priori*, à l'échelle du génome, et donc de découvrir de nouvelles cibles potentielles du FT étudié. Elle est également beaucoup plus puissante que les techniques *in vitro* plus classiques, comme le simple hybride chez la Levure, où l'on peut facilement mesurer de faux-positifs. Elle peut être mise en œuvre chez le peuplier puisque nous disposons d'une séquence de référence, celle de Nisqually-1 (*P. trichocarpa*), sur laquelle peuvent être alignées les petites séquences obtenues.

La technique ChIP consiste à réaliser une liaison covalente *in vivo* entre l'ADN et les protéines en utilisant un agent de pontage chimique ou physique : le pontage doit se faire très rapidement après le prélèvement du xylème pour fixer les FT en place sur la chromatine. La chromatine pontée est purifiée, fragmentée par sonication, puis immunoprécipitée avec des anticorps, reconnaissant spécifiquement le FT que l'on étudie. La technique a été mise au point par mon équipe d'accueil sur cellules cambiales et cellules de jeune xylème en utilisant des anticorps reconnaissant des modifications particulières des histones, spécifiques de l'euchromatine ou de l'hétérochromatine. Dans le cadre de ma thèse, j'ai utilisé cette technique avec des anticorps anti-MYB090 et anti-MYB221 sur des échantillons de jeune xylème de bois de tension (2 répétitions biologiques par anticorps). Pour MYB221, 57 cibles potentielles ont été identifiées ; sur 5 cibles choisies, 4 ont été confirmées par ChIP-qPCR sur un troisième échantillon biologique indépendant (un FT MYB, un FT de type bZIP, une shikimate déshydrogénase et une O-acyl transférase membranaire). Pour MYB090, 874 cibles potentielles ont été identifiées : des FT, des gènes impliqués dans la synthèse des composés pariétaux, telles des cellulose synthases, glycosyltransférases, protéines à arabinogalactanes à domaine fascicline, F5H, COMT.

4.3- Etude fonctionnelle des FT dans des peupliers transgéniques

La transgénèse est une technique permettant de modifier l'expression d'un gène *in planta*, et ainsi d'établir un lien de causalité entre modification du génotype et modification du phénotype. C'est donc une technique de choix pour démontrer la fonction d'un gène. Dans le cas particulier de l'étude de FT, nous avons choisi de faire de la surexpression avec un promoteur constitutif fort, plutôt que de l'extinction de gène car nous avons craint de ne

pas voir d'effet phénotypique, de par la possible redondance fonctionnelle des FT. Nous avons choisi deux approches, soit la surexpression du FT natif, soit la surexpression du FT fusionné à un motif répresseur de type SRDX (Hiratsu *et al.* 2003). Ce FT modifié entre alors en compétition avec le FT endogène et entraîne une répression systématique des gènes cibles du FT endogène. On obtient alors un phénomène de répression dominante et théoriquement, un phénotype correspondant à une extinction du FT. Cette stratégie est particulièrement adaptée aux FT activateurs de transcription, nous l'avons cependant utilisée pour nos 3 FT même si MYB156 et MYB221 sont susceptibles d'être des répresseurs potentiels, car orthologues d'EgMYB1. La surexpression de MYB090 a entraîné une diminution importante de la croissance et de la teneur en lignines des tiges des peupliers transgéniques. Au niveau histologique, le xylème se caractérise par une absence nette de lignification dans les rayons. Au contraire, si la surexpression de la forme MYB090-SRDX a fortement affecté la croissance des plantes, nous n'avons pas observé de modifications notables au niveau histologique ou biochimique. La surexpression de MYB156 a entraîné une baisse très importante de la lignification des parois cellulaires dans le xylème, à l'exception de celles des vaisseaux, associée à une diminution de croissance. Les plantes répressives dominantes de MYB156 sont naines mais avec des racines très développées. La surexpression de MYB221 n'a pas entraîné de modification notable de la lignification du xylème secondaire. Par contre, les plantes sur-exprimant la forme MYB221-SRDX sont affectées au niveau de leur croissance et de la lignification des parois. Ces résultats suggèrent que les trois FT étudiés jouent un rôle dans la régulation de la lignification, le rôle de chacun sera discuté au regard des résultats obtenus dans les 3 chapitres de la thèse.

Matériel et Méthodes

Tableau (a) : Nombre de lignées acclimatées pour chaque transformation

vague	date	717-1B4	MYB221 OE	MYB221- SRDX OE	MYB156 OE	MYB156- SRDX OE	MYB090 OE	MYB090- SRDX OE
1	26-04-12	5 plants	11 lignées (22 plants)	10 lignées (20 plants)	0	6 lignées (12 plants)	4 lignées (8 plants)	0
2	06-06-12	2 plants	13 lignées (26 plants)	14 lignées (28 plants)	18 lignées (36 plants)	4 lignées (8 plants)	9 lignées (18 plants)	0
3	03-07-12	5 plants	0	0	3 lignées (6 plants)	0	7 lignées (14 plants)	0
4	04-09-12	4 plants	0	0	0	0	0	4 lignées (8 plants)

1- Matériel végétal et échantillonnage

1.1- Plantes destinées à la ChIP

Pour les expériences de ChIP, nous avons utilisé le clone 101-74 parce que c'est un *Populus trichocarpa*, dont la séquence génomique est très proche de la séquence de référence (Nisqually-1, *P. trichocarpa*). Les plants issus de boutures ligneuses ont été cultivés en pot de 4L pendant 6 mois en serre jusqu'à atteindre une hauteur d'environ 2 m et un diamètre de plus d'1 cm à la base de la tige. Les plants sont alors mis en inclinaison pendant 7 jours pour induire la formation de bois de tension. Les tiges sont prélevées sur environ 1 m à partir de la base, par lot de 6 à 9 plants pour constituer un échantillon biologique. Les parties apicales de plus faible diamètre ne sont pas prélevées. A la base des tiges, des fragments de 5 cm sont coupés, séparés à la trancheuse en face supérieure (TW) et inférieure (OW), puis écorcés; le jeune xylème de chaque demi-tige est gratté légèrement au scalpel, immédiatement broyé dans l'azote liquide, et congelé à -80°C avant extraction d'ARN (voir 2-). Le reste des tiges est également coupé en fragments de 5 cm de longueur, puis séparé en TW et OW avant d'être utilisés pour les expériences de ChIP (voir 9-).

1.2- Lignées transgéniques

Les lignées transgéniques ont été produites par transformation génétique du clone INRA 717-1B4 (clone femelle hybride de *Populus tremula* x *P. alba*) (voir 6-). Pour le premier criblage, au moins deux plantes par lignée transgénique ont été acclimatées en serre en 4 vagues d'acclimatation en 2012 (Tab.a) en pot de 2L. Pour chaque lignée, des mesures de hauteur, diamètre de la tige (au-dessus de la première feuille, à environ 2 cm de la base), du nombre de branches sylleptiques et d'autres observations phénotypiques ont été réalisées après 5, 8 et 10 semaines de culture. Les plants ont été échantillonnés à 12 semaines. Chaque tige a été prélevée de la base jusque la première feuille mature. Les feuilles sont éliminées. Un tronçon de 3 cm, destiné aux études histologiques (voir 10-), est prélevé à la base de chaque plant et stocké à -20°C. Un tronçon correspondant aux 10 cm suivants, destiné aux extractions d'ARNs totaux (voir 2-) est découpé en plusieurs morceaux pour être immédiatement stocké dans l'azote liquide, puis à -80°C.

En 2013, un nombre plus limité de lignées transgéniques choisies à partir du premier criblage de 2012 a été acclimaté en serre pour une période de 10 semaines. Toute la partie aérienne (tige avec feuilles) a été coupée puis pesée pour connaître la biomasse aérienne produite. Ensuite, la tige a été effeuillée; un tronçon de 3 cm à la base de la tige a été prélevé pour les études histologiques (voir 10-). Le reste est écorcé puis séché et broyé pour les analyses biochimiques (voir 11-).

2- Extraction d'ARNs totaux et synthèse d'ADNc

Les ARNs totaux sont extraits à partir de 200 mg environ de poudre de xylème broyé selon le protocole modifié de Chang *et al.* (1993), puis purifiés avec le kit RNeasy MinElute (Qiagen). La transcription réverse est réalisée sur 2 µg d'ARN totaux avec une Superscript III (Invitrogen) en présence d'oligos d(T).

3- Extraction d'ADN génomique

L'ADN génomique (ADNg) des plantes transgéniques a été extrait à partir d'une jeune feuille fraîche prélevée au niveau de l'apex. Ces feuilles sont congelées dans de l'azote liquide puis broyées à froid. L'extraction a été faite en plaque 96 puits selon le protocole du kit NucleoSpin® 96 Plant II de Macherey-Nagel avec quelques modifications : le tampon PL1 a été utilisé pour la lyse cellulaire pendant 1h à 65°C. La vitesse de centrifugation choisie est de 5600 g. L'ADN est élué dans un volume final de 80 µl.

4- PCR

La PCR a été utilisée sur l'ADNg des lignées transgéniques et du WT (dilué au 1/100) pour détecter la présence du gène de résistance à l'hygromycine, et sur des ADNc issus de jeune xylème de lignées transgéniques, WT ou 101-74 (dilué au 1/4) pour détecter le niveau d'expression des FT. La réaction a été faite dans un volume total de 25 µl avec une concentration des amorces de 200 nM. Une première dénaturation a été réalisée pendant 10 min à 94°C, ensuite 29 à 35 cycles comprenant une étape de dénaturation à 94°C pendant 45 s, une étape d'hybridation à 55-60 °C (selon les amorces) pendant 45 s et une étape

d'élongation à 72 °C pendant 1 à 4 min (selon la taille de l'amplicon). Enfin la réaction se termine par une élongation finale à 72 °C pendant 10 min. Les amplicons sont examinés sur gel d'agarose 1,5%, après coloration au bromure d'éthidium.

5- Q-RT-PCR

Pour les études d'expression quantitative, des amorces gène-spécifiques sont dessinées à l'aide de QuantPrime (<http://www.quantprime.de/>) pour un T_m de 60 °C et une taille d'amplicon idéale comprise entre 100 et 150 bp. Chaque couple d'amorces a été testé *in silico* à l'aide du site web <http://gramene.org/Multi/blastview> et <http://popgenie.org/> afin de vérifier la spécificité de celles-ci. La réaction se fait dans un volume réactionnel de 12 µl avec le kit Platinum SYBR Green superMix à partir d'ADNc dilué au 1/4^{ème}. 200 nM d'amorces sens et anti-sens ont été utilisées. Après une étape de chauffage à 50°C pendant 2 min puis 90°C pendant 2 min, les conditions de chaque cycle PCR (40 cycles au total) consistent en : une étape de dénaturation (95°C durant 15 s), une phase d'hybridation (60°C pendant 30 s) et une étape d'élongation (68°C durant 20 s) avec une mesure de la fluorescence. Le niveau d'expression relative (Qr) a été quantifié selon la méthode de ΔC_t en utilisant la formule suivante :

$$Qr = \frac{E^{\Delta C_t \text{ cible (témoin - échantillon)}}}{E^{\Delta C_t \text{ réf (témoin - échantillon)}}$$

6- Production des plantes transgéniques

6.1- Réalisation des constructions

Les séquences codantes de MYB090, MYB156 et MYB221 ont été clonées par PCR à partir d'une matrice d'ADNc préparée à partir d'ARN extraits de jeune xylème de bois de tension de 717-1B4 (*Populus tremula x P. alba*). Les amorces ont été dessinées d'après les séquences des transcrits identifiés chez Nisqually (*P. trichocarpa*), respectivement Potri.015G033600.1, Potri009G134000.1 et Potri004G174400. Les produits PCR ainsi

Tableau (b) : Amorces utilisées pour le clonage des FT avec motif SRDX ou non dans le vecteur d'entrée

Nom de l'amorce	Séquence
MYB221&156_attB1_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGGGAAGGTCTCCTTGCTG
MYB090_attB1_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCATGTGTACTAGAGGCCATTG
attB2_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGCGAAACCCAAACGGA
SRDX_attB2_R	TTAAGCGAAACCCAAACGGAGTTCTAGATCCAGATCGAGCCCACATTCATCTCCAAACCTCTAT

Tableau (c) : Amorces utilisées pour fusionner le motif SRDX aux FT en 3'

Nom de l'amorce	Séquence
MYB221_SRDX_F	ATGTGTACTAGAGGCCATTGGA
MYB221_SRDX_R	ttaagcgaacccaaacggagttctagatccagatcgagcccACAAGCTGAGAAAAAATCAAAGA
MYB156_SRDX_F	ATGGGAAGGTCTCCTTGCT
MYB156_SRDX_R	ttaagcgaacccaaacggagttctagatccagatcgagcccATATTTTCATCTCTAAACTTCTAT
MYB090_SRDX_F	ATGTGTACTAGAGGCCATTGGAG
MYB090_SRDX_R	ttaagcgaacccaaacggagttctagatccagatcgagcccACAAGCTGAGAAAAAATCAAAGAAT

Tableau (d) : Amorces utilisées pour l'étude d'expression du transgène

Nom de l'amorce	Séquence
MYB221_CDS_185F	TTTGTTTTGCTTGCAGCTTG
MYB156_CDS_184F	TTGTTTTGCTTGCAGTTTGG
MYB090_CDS_160F	CCAAGTAAACACGGACTCCAA
SRDX_184/185/160R	TTCTAGATCCAGATCGAGCCC
MYB221-F	TCAGCACTACTACCTCAGTTAAAGAA
MYB221-R	CAGTTTTGGAGCCAGCACTAC
MYB156-F	CTTTCACCACGACTACTACTTCAG
MYB156-R	TTGTGCTAGTGCTGTCACTG
MYB090-F	GGCCTGCAGAAGATGAGAAG
MYB090-R	CGTTTACCAAAGCTTGAGCAG

amplifiés ont été clonés dans le vecteur pCR2.1 (Annexe 2) par TOPO/TA cloning selon les instructions du fabricant (Invitrogen). L'identité des clones a été vérifiée par séquençage.

Différentes constructions ont ensuite été préparées:

- Les séquences codantes complètes des 3 FT ont été amplifiées par PCR à partir des vecteurs pCR2.1 avec des amorces de type Gateway® portant les sites attB1 / attB2 (tab.b), et ont été introduites par recombinaison de type BP dans le vecteur d'entrée pDONRZéo (Invitrogen™) (Annexe 3)
- Le motif SRDX a été fusionné aux séquences des FT par PCR en utilisant des amorces adéquates (Hiratsu *et al.* 2003) (tab.c). Les FT modifiés sont clonés dans pCR2.1 par TOPO/TA cloning (Invitrogen). Les séquences codantes des 3 FT-SRDX ont été amplifiées par PCR à partir des vecteurs pCR2.1 avec des amorces de type Gateway® portant les sites attB1 / attB2, et ont été introduites par recombinaison de type BP dans le vecteur d'entrée pDONRZéo (Invitrogen™) (tab.b).
- Les FT ou les FT-SRDX dans pDONRZéo ont été transférés par recombinaison de type LR dans le vecteur binaire pMDC32 (Curtis et Grossniklaus, 2003) (Annexe 4) permettant une expression dans la plante de façon constitutive (promoteur 35S). Le gène de sélection est le gène de résistance à l'hygromycine.

Les séquences réalisées aux différentes étapes de clonage ont été analysées en faisant des alignements par rapport à la séquence théorique, à l'aide du logiciel GenalysWin3.4.19a et/ou NCBI Blast (bl2seq) et/ou ClustalW2.

6.2- Transformation génétique

Les plasmides extraits avec le kit NucleoSpin (Macherey-Nagel) selon les recommandations du fabricant ont été vérifiés par PCR, par restriction et par séquençage, puis introduits par électroporation dans la souche d'*Agrobacterium tumefaciens* C58 (pMP90). La transformation génétique a été réalisée par coculture avec des explants de tiges blessées de 717- 1B4 selon le protocole de Leplé *et al.* (1992) modifié. Les modifications suivantes ont été apportées à ce protocole afin d'augmenter le nombre des cals transformés: après la

Tableau (f) : Amorces utilisées pour l'étude d'expression de certaines cibles de MYB090 chez les plantes transgéniques

Nom de l'amorce	Séquence	Taille de l'amplicon
Potri.006G251900F	CCTGCTGTTTGCCTTGTCAGT	203
Potri.006G251900R	AAGACGGCAAAGAGATGAGCTG	
Potri.005G058300F	GAGCACTTGACCAGGGAGAG	208
Potri.005G058300R	CCGGCAGAATCCAGAGGAAG	
Potri.019G120900F	CAGTGGCATCTGCACCAGAAAA	207
Potri.019G120900R	CCGGACACAACCCTACATATATCA	
Potri.002G172100F	CCTGCATGGTTCTAGCTGGG	215
Potri.002G172100R	GTGGTCTGCCAATCCCAGAA	
Potri.012G006400F	AGTAGAGAGTAGGGACCAACCA	216
Potri.012G006400R	TGTGCCTCTTCATCTGATACCT	
Potri.001G406900F	ACCTTAGATGTCTTTCCATCTCCA	215
Potri.001G406900R	CTCCCTCGCTCTCTTGAATGA	
Potri.016G098400F	ATCTAGGGCCAGAGATGTGGA	200
Potri.016G098400R	ATTGCACCCAAAAACGGGGA	
Potri.002G059900F	CCTAGAAACAGCAAATGTTGCGA	203
Potri.002G059900R	TGACCTGAGGAGAGACGCTT	

Tableau (g) : Amorces utilisées pour l'étude d'expression de certains gènes de biosynthèse des lignines, cellulose et xyloxygènes chez les plantes transgéniques

Nom d'amorce	séquence	taille de l'amplicon
CAD4-F	TCGATACTGTGCCTGTGGTTCAC	146
CAD4-R	CCCGGTGATTGACTTTCTCCAAG	
CesA-F	TGCACGCTGGTAGTTGAAGAGGAG	175
CesA-R	CCCATCTTGCATGATCCACCCATC	
IRX8-F	AAGGTTAGAGCAAGAAGGGTTTCG	141
IRX8-R	CAGGCAACCTAGTGTAGAGCAC	
IRX9-F	ACAGATCCCGAAGCAGAGTTGG	165
IRX9-R	CATTGGCCATGTTCCAAACACCTC	
PAL1-F	TGTCTTTGCTTACGCCGATGACC	190
PAL1-R	TGCGCTCTCAACCTCTTTAGGC	
CESA4-F	ACCTTGCATGCCAAAGATGAGC	137
CESA4-R	GGAAAGCCACACACATGACAAGC	
CESA7-F	ACTGGTGTGAGATCAAGACCTGTG	141
CESA7-R	TTTCTTTGCGTCCCATCTTGCG	
CESA8-F	TCGACAGGCTATCTGCAAGGTATG	144
CESA8-F	AGTCCACAGCAAGGATGGAAAGG	
CCRF	CATGTTAGGGATGTGGCACTAGC	165
CCRR	CTTCTCATCTGAGCACTTGGTAGG	

décontamination, les explants ont été mis directement sur un milieu MS2 + antibiotique de sélection (hygromycine) et la sélection a été faite pendant 21 jours en continu à l'obscurité avant le passage à la lumière.

6.3- Caractérisation moléculaire des plantes transgéniques

La transformation des plantes a été vérifiée au niveau moléculaire par PCR sur ADNg en utilisant des amorces permettant de détecter la présence du gène de résistance à l'hygromycine *hpt* (HPT421_F : AACCACGGCCTCCAGAAGAAG; HPT421_R : ACCTGCCTGAAACCGAACTGCCCCG). Pour ceci, l'ADN génomique a été dilué au 1/100 et 5 µl de matrice ont été utilisés pour la réaction d'amplification de 35 cycles et une température d'hybridation de 61°C. La sur-expression du transgène a également été vérifiée par RT-PCR, à l'aide des amorces présentées dans le tableau (d).

Les plantes transgéniques ont été également caractérisées pour l'expression de gènes cibles de MYB221 et de MYB090 par RT-qPCR (les amorces sont présentées dans les tableaux (e) et (f)), ainsi que pour l'expression de gènes de biosynthèse des lignines, cellulose et xylanes (amorces présentées dans le tableau g).

Tableau (e) : Amorces utilisées pour l'étude d'expression de certaines cibles de MYB221 chez les plantes transgéniques

Nom de l'amorce	Séquence	Taille de l'amplicon
Potri.013G000600F	CTCTCCAATCCAATCGATGCC	203 pb
Potri.013G000600R	TCGAGAGAGAGTTGAAGAGATCG	
Potri.010G137000F	CAACCCGGTAACGGATTCCT	201 pb
Potri.010G137000R	CCGTCTTCAACTTGTGGGC	
Potri.T011400F	TCTTGCAATCTTCTCCCAGCA	201 pb
Potri.T011400R	TCTTCTTCTTTAGTCCAGGCACC	
Potri.013G029800F	AGGCTGGAAGCACCATTGTT	214 pb
Potri.013G029800R	TGGACATGACTTGGCTGCAA	
Potri.006G251800F	TATCAGCCCAGCAAGCAAGG	220 pb
Potri.006G251800R	ACAATGTCATAGCCTCCTGGTA	
Potri.004G067000F	ACCATCTCGTTTGCGAAACCT	220 pb
Potri.004G067000R	GGCAGAGAGAGGTAGTGACCA	

Tableau (h) : Amorces utilisées pour le clonage des 3 FT dans pSC-A

<i>Nom de l'amorce</i>	<i>Séquence</i>
MYB221+156_SpeI_F	ACTAGTATGGGAAGGTCTCCTTGCTG
MYB156_SpeI_R	ACTAGTATATTTTCATCTCTAAACTTCTA
MYB221_SpeI_R	ACTAGTACATTTTCATCTCCAAACCTC
MYB090_SpeI_F	ACTAGTATGTGTACTAGAGGCCATTG
MYB090_SpeI_R	ACTAGTACAAGCTGAGAAAAAATCAA

7 - Obtention des anticorps polyclonaux anti- MYB90 et anti-MYB221

Les anticorps polyclonaux anti-MYB90 et anti-MYB221 ont été produits chez le lapin par la société Agro-Bio (La Ferté Saint Aubin) à partir de protéine recombinante produite dans *E. coli* et purifiée par affinité pour son tag 6HIS. Nous leur avons également confié la purification des anticorps par affinité sur antigène. Ainsi, après couplage de 5 mg de MYB90 recombinant sur gel CNBr sepharose B, 1,8 mL d'anticorps purifié anti-MYB90 à 0,45 mg/mL a été obtenu. Après couplage de 3,4 mg de MYB221 recombinant sur une colonne Hi-trap NHS, nous avons également obtenu 1,8 mL d'anticorps purifié anti-MYB221 à 0,5 mg/mL. Les anticorps purifiés nous ont été fournis en solution tampon PBS 0,15M pH7,4 + 0,09% azide de sodium et ont été conservés à -20°C.

8 - Localisation cellulaire des FT en système cellulaire hétérologue

Les FT ont été clonés par restriction-ligation au site SpeI dans le plasmide pSC-A (Annexe 1), permettant l'expression du FT fusionné à la protéine fluorescente YFP (Yellow Fluorescent Protein) sous le contrôle du promoteur 35S. Pour réaliser ce clonage, le FT a été préalablement amplifié par PCR avec des amorces gène-spécifiques permettant d'introduire les sites de restriction SpeI (tab. h), cloné de façon transitoire par TOPO-TA cloning dans pCR2.1. Le fragment correspondant au FT a été excisé par SpeI, puis ligué dans pSC-A ; le sens d'insertion du FT a été vérifié par restriction. Les différents plasmides produits ont été purifiés par miniprep (Kit Nucleospin Plasmid de Macherey–Nagel). Des tapis cellulaires de *Catharantus roseus* ont été préparés par filtration d'une suspension cellulaire sur un disque de papier filtre. Le tapis est ensuite déposé sur un milieu de culture solide (Gamborg B5, ANA 10µM, agar 8g/l, céfotaxime 200µg/ml) puis incubé en chambre de culture pendant 48h. Pour chaque tir, 500 ng de plasmides (« FT-YFP » = FT couplé à la YFP) et 200 ng de plasmide-marqueur («nucleus-CFP» = marqueur du noyau couplé à la CFP) sont fixés sur 500µg de billes. Les tirs sont réalisés avec le système Bio-Rad PDS1000/He, sous vide à 28inch Hg et avec des disques de rupture de 1100 psi. Le tapis cellulaire est disposé à 6 cm du dispositif de tir. Après transformation, les cellules sont incubées 15-20h en chambre de culture. Les cellules transformées sont ensuite observées sous microscope à épifluorescence Olympus.

9- ChIP-Seq et ChIP-PCR

9.1- Pontage de la chromatine *in vivo*

La fixation, dans les cellules vivantes, des protéines entre elles et des protéines aux acides nucléiques se fait par un traitement au formaldéhyde. Cet agent chimique de liaison permet de fixer la chromatine dans l'état de régulation transcriptionnelle dans lequel elle se trouvait au moment du prélèvement. L'avantage de l'utilisation du formaldéhyde est que la réaction est totalement réversible à la chaleur.

Les tronçons de tige de peuplier sont fendus en 2 au massicot pour séparer bois tendu et bois opposé. Après écorçage, les échantillons sont immédiatement immergés dans le tampon de pontage (Tris HCl 10mM pH8, saccharose 0,4M, MgCl₂ 10mM, sodium sulfite 20mM, L-cystéine 20mM, PVP40 2%, β-ME 5mM, Pluronic F-68 0,01%, PMSF 1mM) contenant 1% de formaldéhyde. Le pontage se fait par infiltration sous vide pendant 20 min à - 900 mbar. Le pontage est arrêté par ajout de glycine (125mM final) dans le tampon de pontage ; l'infiltration sous vide est poursuivie dans les mêmes conditions pendant 5 min. Les échantillons sont ensuite rincés plusieurs fois dans de l'eau à 4°C. Le jeune xylème est gratté très légèrement au scalpel, broyé dans l'azote liquide et conservé également à -80°C.

9.2- Extraction et sonication de la chromatine

Nous avons ensuite préparé les noyaux selon le protocole mis au point par l'équipe de Werner Aufsatz en Autriche sur *Arabidopsis* (http://epigenesys.eu/images/stories/protocols/pdf/20111025150640_p13.pdf). 1.5 g de poudre de bois congelée sont soumis à 3 cycles d'extractions/centrifugation avec des tampons de concentration croissante en saccharose. Le culot de noyaux est repris dans 300μl de tampon de lyse (Tris HCl 50mM pH8, EDTA 10mM, protease inhibitor cocktail for plant cell Sigma 0,2% v/v, SDS 1%, β-ME 14mM) permettant d'extraire la chromatine. Après macération pendant 1 heure à 4°C sous agitation, la chromatine est clivée par ultrasons en fragments de 500 à 1000 pb, dans un Bioruptor® Diagenode pendant 30 min avec des pulses de 30s. Après centrifugation 10 min à 12000 g et à 4°C, le surnageant contenant la chromatine est récupéré et conservé à -80°C en attendant d'être utilisé en ChIP.

Tableau (i) : Amorces utilisées pour déterminer l'enrichissement de l'ADN immunoprécipité en certaines cibles de MYB221

Nom de l'amorce	Séquence	Taille de l'amplicon
ChIP_MYB221_25F	TGTTAACTTGAAATCCTTGTCCT	162pb
ChIP_MYB221_25R	GTGGAGCAATTAATCCCCAAA	
ChIP_MYB221_Out_25F	TTCAATTGCCTTGCTTGCCC	200 pb
ChIP_MYB221_Out_25R	CATCGTAAAGGTGTGTGCGG	
ChIP_MYB221_76F	CCGGACACCCACATTAAACTAAA	191 pb
ChIP_MYB221_76R	ACTTGAAGAGTTGTTGGAAGCA	
ChIP_MYB221_Out_76F	GGCTAAATCATGTGGGGAAGC	224 pb
ChIP_MYB221_Out_76R	ATTACATGGAAAAACACTGCAACA	
ChIP_MYB221_148F	TCCAATCACAGAAAGAAACACTACA	181 pb
ChIP_MYB221_148R	TGGGATTTGACTAAGTTGGTGAG	
ChIP_MYB221_Out_148F	ACTGCTGCTGATAGCTAAACATAA	218 pb
ChIP_MYB221_Out_148R	GCACGAGACGAGAGACACAA	
ChIP_MYB221_309F	AGAAGAAGGTGGCCAAGTAGT	197 pb
ChIP_MYB221_309R	GTTGCTCCAACAGGCTAACG	
ChIP_MYB221_Out_309F	GTTTCCCTGCCAGTGATCGT	206 pb
ChIP_MYB221_Out_309R	AGCCAGAGAGGTTGGCTCTA	
ChIP_MYB221_353F	GTCTGTCCCTAACACCAACCC	138 pb
ChIP_MYB221_353R	AATTATGTTAGGCAGGGGCGG	
ChIP_MYB221_Out_353F	GGGTCACGGATCCTTCAACAT	216 pb
ChIP_MYB221_Out_353R	CTCATGCCACAAGGTCGAGA	
ChIP_MYB221_368F	CAGTGTTTTCCCCACGCCT	184 pb
ChIP_MYB221_368R	TCAATGAGAGTTTGCGTTTCTCTC	
ChIP_MYB221_Out_368F	ACTCAAGCTTAGAACAATGGACTT	198 pb
ChIP_MYB221_Out_368R	AAAAATAACCCGCGGCATCG	

Tableau (j) : Amorces utilisées pour déterminer l'enrichissement de l'ADN immunoprécipité en certaines cibles de MYB090

Nom de l'amorce	Séquence	Taille de l'amplicon
ChIP_MYB090_135F	AGACACTGATACGTGTCGCA	195pb
ChIP_MYB090_135R	CTCAAGCTGGAGCCGAAGAA	
ChIP_MYB090_Out_135F	ACTCTGCTTACACCTCCCTCT	185pb
ChIP_MYB090_Out_135R	AAAATTGGCATGGAGCCCAG	
ChIP_MYB090_436F	CACGAAGGCCAAAACATCGG	196 pb
ChIP_MYB090_436R	TTTTCGGGATGGAAGGAGGC	
ChIP_MYB090_Out_436F	AGTTATCACCAAACATTTCTTTCCT	207 pb
ChIP_MYB090_Out_436R	TCGAAATACTACGGTACTTGAC	
ChIP_MYB090_789F	TGCTAGAAACAATGCAAAAATGCT	191 pb
ChIP_MYB090_789R	GACCCTCCACCTTTTGAGT	
ChIP_MYB090_Out_789F	ACCCTAGCGTTAACTTGGA	218 pb
ChIP_MYB090_Out_789R	AGGAAAATTTGATGTTGGTCTCA	
ChIP_MYB090_5F	TCGGCAGCGATTTGTGTAGT	217 pb
ChIP_MYB090_5R	ATGGCAACACGGGTACTAGC	
ChIP_MYB090_Out_5F	AATTAATTTTATTGTCGCCCCGA	183 pb
ChIP_MYB090_Out_5R	TGGAGTGAAACAAGGATTAAATGG	
ChIP_MYB090_23F	TCGAGAAAAAGAAGAAGCACAACA	194 pb
ChIP_MYB090_23R	ACCCAAGTCTCTACGTCAATGT	
ChIP_MYB090_Out_23F	CTGCTGCAGGCTAGATCAGA	373 pb
ChIP_MYB090_Out_23R	AACTCACCAGTGTGCGCAAA	
ChIP_MYB090_3F	AGGACTGGGATTTGTGGGGA	180 pb
ChIP_MYB090_3R	GCATGTGCCAAGACACCTTT	
ChIP_MYB090_Out_3F	ATTGTTGATGTATCAATTAAGCGT	184 pb
ChIP_MYB090_Out_3R	TTGTGAGTGTGGTAGTGATTGCT	
ChIP_MYB090_95F	ATTGCAGAAACGGCTTCCAT	200 pb
ChIP_MYB090_95R	AGAAGCCACGTGTATAGGAACT	
ChIP_MYB090_Out_95F	AGGCCTAGTTAAACAATCCTGAC	207 pb
ChIP_MYB090_Out_95R	TGTGGTACTAGGTTGAGATTGA	
ChIP_MYB090_451F	TCAAAATACGGCTCTACACGC	180 pb
ChIP_MYB090_451R	TAAACCAACAGCCACCGTCT	
ChIP_MYB090_Out_451F	TCACTTCAACCAACCAATACATGAC	209 pb
ChIP_MYB090_Out_451R	GACCGTTCAAGGGGAGGAAA	

9.3- ChIP et évaluation de l'enrichissement en cibles par PCR

Nous avons réalisé l'immunoprécipitation de chromatine avec les kits HighCell#ChIP et IPure de Diagenode. Dans un premier temps, l'anticorps dirigé contre la protéine d'intérêt est lié à des billes magnétiques couplées à la protéine A. On incube ensuite ces billes avec la chromatine soniquée. Après plusieurs lavages, on isole ensuite le complexe billes / protéine A / anticorps / FT / ADN grâce à un aimant. On libère ensuite l'ADN du complexe par chauffage à 65°C pendant 4 heures. L'ADN est purifié grâce aux billes magnétiques du kit IPure (Diagenode). Un échantillon Input est traité de la même façon sauf qu'il n'est pas mis à incuber avec les billes / protéine A / anticorps.

L'ADN immunoprécipité (IP) et l'Input peuvent être analysés par qPCR ou être séquencés. Les amorces utilisées pour déterminer par qPCR l'enrichissement de l'ADN immunoprécipité en certaines cibles de MYB221 et MYB090 sont présentées dans les tableaux (i) et (j) respectivement. La réaction de qPCR est réalisée avec le kit Platinum SybrGreen qPCR super mix (Invitrogen) dans un volume réactionnel de 12 µl à partir de 5 µl de matrice ADN IP ou d'Input dilué au 1/100^e en présence de 200nM de primers. La quantification relative de l'ADN a été calculée par rapport à une gamme de calibration établie à partir du produit PCR de la cible étudiée amplifiée à partir d'ADN génomique de 101-74. Le calcul de l'enrichissement (% Input) est le pourcentage de cibles IP par rapport aux cibles présentes dans l'Input.

9.4- Séquençage et analyses bioinformatiques de l'ADN immunoprécipité

Le séquençage a été réalisé sur la plateforme de séquençage de l'IGBMC de Strasbourg, Institut de Génétique et de Biologie Moléculaire et Cellulaire (<http://www-microarrays.u-strasbg.fr/index.php>) avec la technologie HiSeq2500 d'Illumina. La qualité et la quantité des ADN enrichis fournis sont analysées sur un fluorimètre (Qubit, Invitrogen). Les librairies sont préparées à partir des fragments de 200pb des ADN enrichis, auxquels on ligue des adaptateurs de séquençage possédant un index (code barre de ≥ 6 nt, permettant de multiplexer jusqu'à 3 échantillons par flowcell). Le profil de la librairie est vérifié par électrophorèse capillaire (Bioanalyzer, Agilent). Le séquençage est simple (single read) et génère des tags de 50 bases (fig M.1). Les fichiers de séquences 'brutes' sont au format

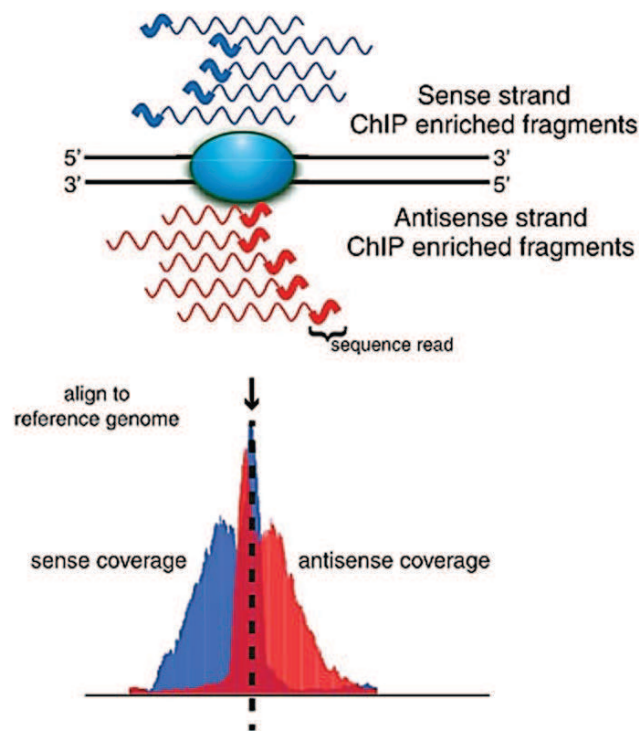


Figure M.1 : Schéma illustrant le principe du séquençage haut débit de l'ADN immunoprécipité.

Les fragments d'ADN immunoprécipité (ligne noire caractère gras) sont séquencés en 5' sur leurs 50 premières bases de chaque brin d'ADN, appelés "tags" (lignes rouge et bleue en caractère Gras). Ces "tags" sont alignés sur le génome de référence. L'alignement des tags génère 2 pics (un pic sens et un pic antisens respectivement en bleu et rouge), de part et d'autre du site de fixation du FT d'intérêt. Le site d'enrichissement correspondant à la zone de fixation de FT à l'ADN est indiqué par une flèche (schéma issu de Wilbanks *et al.* 2012).

FASTQ (fasta+qualité). Ensuite, ces tags sont alignés sur le génome de référence du peuplier. On obtient des fichiers BAM et BED qui sont ensuite utilisés par le logiciel CisGenome. L'alignement des tags va générer 2 pics, 1 pour chaque brin d'ADN, de part et d'autre du site de fixation du facteur de transcription. On observe une distribution bimodale des tags avec la présence de 2 pics (fig M.1). Ces 2 pics correspondent à un enrichissement de l'alignement des tags dans ces zones du génome. Les régions génomiques correspondant à ces deux pics encadrent la zone protégée de la sonication par la présence du FT. L'échantillon Input est également séquençé en suivant la même démarche. Les séquences des tags obtenus sont dispersées tout au long de la séquence du génome de peuplier et sont utilisées comme contrôle pour éliminer le bruit de fond.

Le logiciel CisGenome v2 (Hongkai *et al.* 2008) permet l'analyse de ces tags et la génération des sites d'enrichissement. Les tags issus de l'échantillon IP et du contrôle Input sont repositionnés sur la séquence du génome à l'aide des fichiers .BED. Le logiciel exploite la distribution bimodale des tags pour les recentrer et générer un seul pic (fig M.1). L'outil compare la distribution des pics le long du génome dans l'échantillon IP avec celle de l'Input. Il ne garde au final que les pics enrichis de façon significative par rapport aux séquences ADN de l'Input.

Les gènes cibles sont recherchés à 3kb de part et d'autre de chaque pic significatif, distance que nous avons choisie comme seuil. Ensuite la position de chaque pic par rapport à la séquence génomique du gène cible a été repérée. Il a été choisi de séparer les pics (à proximité de gènes) en trois catégories. Ainsi, un pic peut être situé dans la région 5' en amont du gène cible, correspondant probablement à une zone promotrice. Il peut être aussi localisé dans la séquence même du gène, nommé "In gene" (généralement dans des introns). Enfin, le pic peut être aussi localisé en aval du gène cible, en 3'. L'outil Biomart du site phytozome (<http://www.phytozome.net/biomart/martview/>) a permis de récupérer l'annotation des gènes se trouvant à proximité des pics (version v3.0). En particulier, les termes de Gene Ontology (GO) qui caractérisent les grandes fonctions d'un gène ont été récupérés, pour tous les gènes ayant un pic en position 5' ou à l'intérieur du gène. Pour classer ces gènes, l'outil en ligne agriGO (Du et al. 2010) a été utilisé pour analyser l'enrichissement spécifique en termes GO. Par la suite, les séquences des pics d'intérêt ont été analysées pour rechercher des motifs

consensus. Cette analyse a été réalisée avec le logiciel MEME-ChIP (Machanick et Bailey 2011).

10- Etudes histologiques

10.1- Coloration Safranine-Bleu Alcian

Cette coloration colore en bleu les parois cellulaires cellulosiques et en rouge les parois cellulaires lignifiées. Des coupes transversales de 30 μ m d'épaisseur de tiges de peuplier ont été réalisées sans inclusion préalable à l'aide d'un microtome rotatif (Leica RM 2155, Leica microsystems, Vienne) équipé de lames jetables. Ces coupes ont été traitées 6 min avec une solution d'hypochlorite de sodium (eau de javel) à 2.4% puis lavées 2 min à l'eau osmosée, déshydratées dans deux bains d'éthanol dilués à 25 et 50% (v : v) puis colorées 3.5 min par de la safranine 1% dans l'éthanol 50% et 6 min par une solution alcoolique de bleu Alcian 1%. Les coupes sont montées entre lame et lamelle avec du baume du Canada et ont été observées sous un microscope optique (Leica DMR, Leica Microsystems, Vienne) équipé d'une caméra numérique (Leica DFC 320).

10.2- Suivi de l'autofluorescence des parois en microscopie confocale

Les tiges de faible diamètre ont été incluses en Historésine Technovit ®7100 (Kulzer, Labonord) après avoir été déshydratées par passage dans des bains d'éthanol de concentration croissante. Les échantillons ont été infiltrés à l'aide d'une solution composée de 1% de durcisseur I (Hardener I) dans la résine Technovit 7100. L'infiltration est réalisée en pilulier par immersion des échantillons déshydratés dans des bains de concentration croissante en résine (20% solution d'infiltration et 80% d'éthanol absolu pendant 2 h; puis 40% solution d'infiltration et 60% d'éthanol absolu pendant 4 h; puis 60% solution d'infiltration et 40% d'éthanol absolu pendant 1 nuit; puis 80% solution d'infiltration et 20% d'éthanol absolu pendant 1 journée; puis 100% solution d'infiltration et passage sous vide modéré pendant 1 journée minimum). Un ml de durcisseur II (Hardner II) est ajouté à 15 ml de solution d'infiltration et le mélange est distribué dans des moules jetables de 400 μ l (7x7x5 mm), avant dépôt des échantillons de façon orientée. La polymérisation est faite 1 heure à température ambiante puis 1 heure à 37 °C. A partir des blocs obtenus, des coupes de 3 μ m ont été réalisées à l'aide d'un microtome rotatif (Leica RM 2155, Leica microsystems,

Vienne) équipé d'un rasoir en carbure de tungstène. Les coupes obtenues ont été montées entre lame et lamelle dans du Fluoromount G.

L'autofluorescence des lignines a été observée au microscope confocal (LSM 700, Zeiss) à une longueur d'onde d'excitation de 488 nm et d'émission de 490 nm avec un filtre LP. Les paramètres de détection de la fluorescence sont conservés entre chaque observation de manière à pouvoir comparer les différents échantillons.

10.3- Détection par immunofluorescence des arabinoxylanes

Pour les analyses d'immunomarquage, les échantillons concernés ont été inclus en résine LR white (London resin company LTD). L'infiltration est réalisée en pilulier par immersion des échantillons déshydratés dans des bains de concentration croissante en résine (20% solution de LR White et 80% d'éthanol absolu pendant 2 h ; puis 40% solution de LR White et 60% d'éthanol absolu pendant 4 h ; puis 60% solution de LR White et 40% d'éthanol absolu pendant 1 nuit ; puis 80% solution de LR White et 20% d'éthanol absolu pendant 1 journée ; puis 90% solution de LR White et 10% d'éthanol absolu pendant 1 journée ; puis solution de LR White 100% passage sous vide modéré pendant une journée minimum). Les échantillons de tiges ont été déposés dans des gélules remplies de LR white. Cette dernière polymérise pendant 24 h à 56 °C. Des coupes de 800 nm ont été réalisées à l'aide d'un ultramicrotome (Ultracut R, Leica) équipé d'un couteau en diamant (histo, Diatome) et déposées sur des lames silanisées.

Avant la réaction d'immunomarquage, les coupes sont réhydratées 10 min avec de l'eau distillée, puis traitées 10 min par un tampon TBS-Tween 0,01%, ensuite 1 heure à l'obscurité par un tampon TBS-Tween 0,01% contenant 3% de BSA. Enfin les coupes sont incubées 10 min avec une solution de dilution TBS-Tween 0,01% - BSA 0,3%. L'anticorps primaire AX1 dilué au 1/100 dans du TBS-Tween 0,01% - BSA 0,3% est déposé sur les coupes et incubé une nuit à 4 °C puis rincé avec du tampon TBS-Tween 0,01% - BSA 0,3%. Un anticorps secondaire anti-souris fluorescent (AlexaFluor633) dilué à 1/500 dans du TBS-Tween 0,01% - BSA 0,3% est déposé sur les coupes et incubé 1 heure à température ambiante. Les lames sont ensuite rincées abondamment par du TBS-Tween 0,01% - BSA 0,3% et enfin par de l'eau distillée. Finalement, les coupes sont séchées, montées entre lame

et lamelle dans du Prolong Gold antifade (Invitrogen). Pour chaque réaction d'immunomarquage, quelques coupes contrôles sont obtenues en omettant l'anticorps primaire. Les marquages fluorescents sont observés au microscope confocal (LSM700, Zeiss) avec une longueur d'onde d'excitation de 630 nm et la fluorescence émise est détectée au-delà de 640 nm. Toutes les images correspondent à des projections d'intensité maximum.

11- Dosage des lignines et de la cellulose

11.1- Extraction

Environ exactement 1 g de poudre de xylème (M1) de chaque échantillon est introduit dans un sachet ANKOM. L'ensemble sachet/poudre est pesé (M2) puis mis à sécher pendant une nuit à 105 °C puis de nouveau pesé après refroidissement dans un dessiccateur (PS1 : poids sec 1). Les sachets sont placés dans un ballon d'extraction par groupe de 8 dans 1 L d'éthanol 95 % porté à ébullition et avec agitation pendant 30 min puis lavés par de l'eau ultra pure chaude. Une deuxième étape d'extraction a lieu dans les mêmes conditions avec de l'eau ultra-pure bouillante pendant 30 min. Ces étapes sont répétées 3 fois ou jusqu'à ce que l'eau de lavage soit incolore. Les sachets contenant la poudre extraite sont placés au four à 105 °C pendant une nuit, puis pesés après refroidissement (PS2 : poids sec 2).

11.2- Dosage des lignines

Pour obtenir le taux de lignines, 300 mg de poudre extraite est hydrolysée dans des bouteilles *Schott* de 250 ml avec 5 mL d'acide sulfurique à 72 %. Les bouteilles sont agitées doucement à température ambiante pendant 2 h avec deux billes en verre par bouteille pour faciliter l'homogénéisation de la pâte. De l'eau ultra-pure est ajoutée pour atteindre une dilution de l'acide sulfurique à 3 % puis les bouteilles sont ensuite autoclavées à 120 °C pendant 1 h. Le mélange autoclavé est ensuite filtré dans des entonnoirs filtrants bien secs et pesés (masse M6). La densité optique (DO) du filtrat dilué au 1/15^{ème} est mesurée à 205 nm à l'aide d'un spectrophotomètre. Cette mesure servira au calcul du taux de lignines solubles (S). Les entonnoirs filtrants contenant la lignine Klason insoluble sont rincés par 500 mL d'eau ultra-pure dans le but d'éliminer tout résidu présent sur l'entonnoir. Les entonnoirs sont séchés une nuit à 105 °C puis refroidis dans un dessiccateur pendant 30 min et pesés (M7)

afin de pouvoir calculer le taux de lignines Klason (K). Le taux de lignines totales (T) correspond à la somme de K et de S.

Ces taux sont calculés à l'aide des formules suivantes :

$$S = [(DO \times Fd \times \text{Volume total}) / ((110 \times PS1)/1000)] \times 100$$

$$K = [(M7 - M6)/PS1] \times 100$$

$$T = K + S$$

(Fd : facteur de dilution)

11.3- Dosage de la cellulose

Pour déterminer le taux de cellulose, environ exactement 1 g de poudre de xylème est pesé (MH) à partir de la poudre de bois précédemment extraite et mis dans une bouteille de masse M1' (avec M2' = MH + M1'). Cette poudre subit une délignification à 76°C pendant 1 heure avec agitation dans 2 mL de solution de chlorite de sodium (NaClO₂) à 27 % et 80 mL de solution tampon acétate (NaOH 0,76 M ; acide acétique glacial CH₃COOH 1,25 M). Ensuite, 2 mL de solution NaClO₂ sont rajoutés chaque heure pendant 3 heures jusqu'à décoloration de la pâte qui correspond à la dégradation de la lignine. La solution pâteuse est versée dans des entonnoirs filtrants pesés (M4'). La pâte qui correspond à l'holocellulose est lavée à l'eau ultra pure à température ambiante jusqu'à décoloration complète. Deux lavages supplémentaires avec 15 mL d'acétone permettent ensuite d'avoir une décoloration totale. Les entonnoirs filtrants contenant la pâte sont séchés une nuit à 105 °C puis refroidis dans un dessiccateur pendant 30 min et pesés (M5'). L'holocellulose sèche est conservée immédiatement dans le dessiccateur. Par la suite, 800 mg environ exactement d'holocellulose sont pesés dans des petites bouteilles de 100 mL (PS3). 5 mL d'une solution NaOH à 17,5 % sont rajoutés et le mélange est agité pendant 5 min à 20 °C à l'aide d'une bille en verre. Cette étape est répétée 2 fois encore toutes les 5 mn et une fois pendant 30 min avec l'ajout de 2,5 mL de solution NaOH 17,5% à chaque fois. 16,5 mL d'eau ultra-pure sont rajoutés dans les bouteilles tout en mélangeant. Le mélange est filtré à l'aide de nouveaux entonnoirs filtrants de masse M6' et lavé par 50 mL de solution NaOH à 8,3 %. L'α-cellulose est lavée 2 fois à l'eau en remplissant les creusets et en filtrant sous vide puis 7,5 mL d'acide acétique à 10 % sont rajoutés pour neutraliser le reste de NaOH. L'α-cellulose est lavée ensuite à l'eau jusqu'à

avoir un pH de 5 à 6. Les entonnoirs filtrants avec l' α -cellulose sont séchés une nuit à 105 °C puis refroidis dans un dessiccateur pendant 30 min et pesés (M7'). Les taux d'holocellulose (H) en % de poudre extraite et d' α -cellulose (A) en % de (H) sont calculés à l'aide des formules suivantes :

$$H = [(M5' - M4') / (MH - (M3' - M2'))] * 100$$

$$A = ((M7' - M6') * H) / (M5')$$

Chapitre 1

Choix des facteurs de transcription

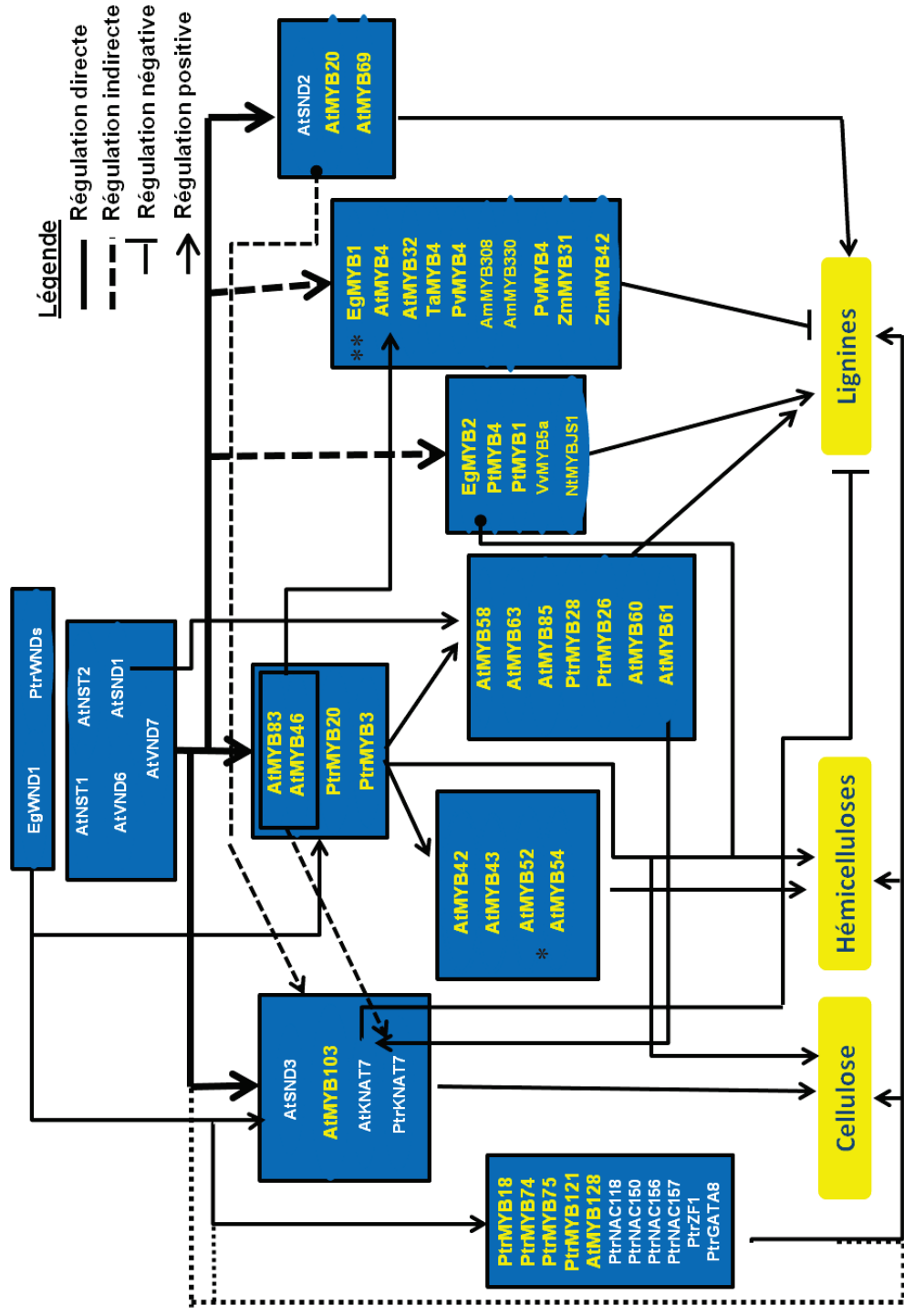


Figure I.1 : Emplacement des 3 FT candidats dans le réseau de régulation de la paroi secondaire. Représentation schématique hypothétique d'une partie du réseau régulationnel de la biosynthèse de cellulose, hémicelluloses et lignines au niveau de la paroi secondaire en se basant sur des FTs appartenant à différentes espèces. L'emplacement de MYB090 est représenté par une (*) et celui de MYB221 et MYB156 par (**). Les lignes continues indiquent une régulation directe potentielle et les lignes discontinues une régulation indirecte potentielle. Les flèches avec pointes aigües (⇒) indiquent une régulation positive et celles avec pointes plates (—) une régulation négative. Les FT de la famille MYB sont présentés en police de couleur jaune.

Pour ce projet de thèse, nous avons besoin de choisir un nombre limité de FT selon différents critères : tout d'abord, nous recherchions des FT exprimés spécifiquement dans le xylème secondaire, et si possible sur-exprimés dans le bois de tension. Nous avons également besoin d'un FT contrôle pour les expériences de ChIP-SEQ, technique permettant de déterminer les cibles *in vivo* de FT : nous souhaitions prendre un FT, impliqué dans le développement du xylème secondaire, et pour lequel au moins une cible avait été démontrée expérimentalement. Ce couple FT / cible connue pouvait ainsi nous servir de référence pour les expériences de ChIP-SEQ, délicates à mettre en œuvre : il pouvait servir à mesurer un enrichissement effectif en cible par PCR après immunoprécipitation de chromatine avant de faire un séquençage haut-débit.

Au démarrage de ce projet de thèse, nous avons donc entrepris une étude bibliographique sur les FT pouvant jouer un rôle dans la formation du bois et/ou du bois de tension (notamment chez le peuplier). Une liste d'environ 40 FT a été établie en croisant données bibliographiques et analyses bioinformatiques. Pour affiner ce choix, nous nous sommes ensuite appuyés sur des données d'expression établies par PCR quantitative à haut débit (technique FLUIDIGM). Pour remplir les critères évoqués précédemment, notre choix s'est porté sur 3 FT : MYB090, un facteur de transcription spécifique du xylème secondaire, dont le gène est sur-exprimé dans le bois de tension, MYB221 et MYB156, 2 FT dont les gènes sont exprimés dans le xylème secondaire, orthologues d'EgMYB1 dont des cibles ont été démontrées expérimentalement chez l'eucalyptus (Legay *et al.* 2007). Nous faisons l'hypothèse qu'au moins un de ces FT a les mêmes cibles qu'EgMYB1. Les 3 FT choisis appartiennent tous les 3 à la famille des facteurs MYB, mais leurs orthologues se situent à plusieurs niveaux du réseau de régulation transcriptionnelle de la formation des différents composants de la paroi secondaire (lignines, xylanes, cellulose) (figI.1).

Dans ce chapitre, nous allons expliciter plus en détail le choix effectué, puis nous décrirons le travail fait pour cloner ces FT, produire des anticorps nécessaires à la mise en œuvre de la ChIP. Nous analyserons les propriétés potentielles de ces FT sur la base de leur séquence et de la bibliographie. Les travaux ont été menés sur 2 clones de peuplier différents. Les FT ont été clonés à partir d'ADN de 717-1B4, clone de *Populus tremula x P. alba*, qui est le clone de peuplier utilisé en routine en transgénèse. C'est ce clone qui sera utilisé dans le chapitre 3 pour les études fonctionnelles *in planta*. Par contre, pour le ChIP-SEQ, les séquences obtenues sont alignées sur un génome de référence (voir chapitre 2). Pour le

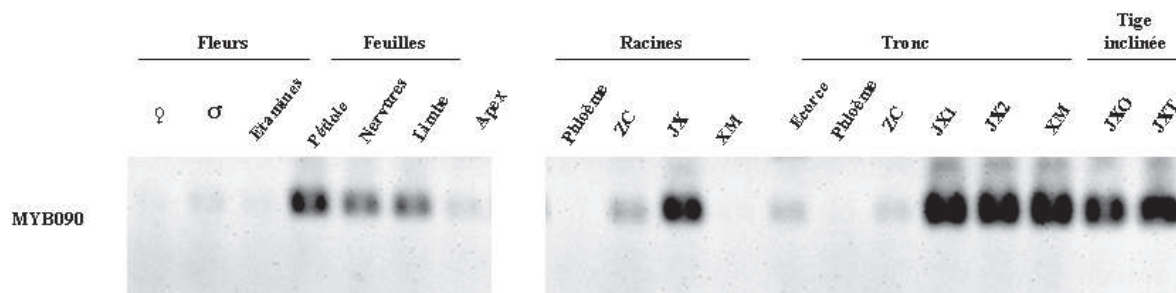


Figure I.2 : Profil d'expression de MYB090 dans différents tissus de peuplier, évalué par RT-PCR semi-quantitative.

Les fleurs mâles et femelles ont été prélevées sur des peupliers *P. nigra* arrivés à maturité sexuelle. Feuilles, racines et tronc ont été échantillonnés sur un arbre de 4 ans (INRA 717-1B4, *P. tremula* x *P. alba*). Les feuilles ont été disséquées au scalpel pour séparer pétiole, limbe et nervures. Le tronc a été écorcé, le phloème a été prélevé à l'intérieur de l'écorce. Côté tige écorcée, les échantillons ont été prélevés séquentiellement, d'abord légèrement au scalpel pour prélever la zone cambiale (ZC), puis au couteau à bois pour prélever le jeune xylème (JX1 et JX2, JX1 étant plus jeune que JX2) puis le xylème mature (XM). De même, les racines lignifiées ont été prélevées et séparées en phloème, zone cambiale (ZC), jeune xylème (JX) et xylème mature (XM). Des apex caulinaires ont été prélevés sur des 717-1B4 de 6 mois élevés en serre. Des 717-1B4 ont également été inclinés en serre pendant 1 mois : le jeune xylème a été prélevé par grattage au scalpel après écorçage côté bois de tension et bois opposé. ZC : zone cambiale, JX : jeune xylème, XM : xylème mature, JXO : jeune xylème de bois opposé, JXT : jeune xylème de bois de tension.

Expression relative de *MYB090* dans le xylème tendu des peupliers 101-74

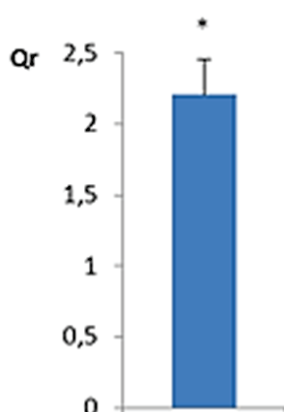


Figure I.3: Expression relative de *MYB090* dans le xylème de bois de tension de peuplier (*P. trichocarpa*, clone 101-74).

L'expression quantitative de *MYB090* (Qr) dans le bois de tension (TW) est rapportée à son expression dans le bois opposé. La barre d'erreur représente l'erreur standard de 4 répétitions. L'astérisque au-dessus du rectangle indique que la valeur de Qr est significative (t-test $p < 0.05$).

moment, la séquence de référence est celle de Nisqually, clone de *P. trichocarpa* (Tuskan *et al.* 2006). Pour le ChIP-SEQ, nous avons donc utilisé le clone 101-74, un clone de *P. trichocarpa*, qui au niveau génétique est plus proche de Nisqually que le 717-1B4. Si les séquences codantes des gènes sont bien conservées entre *P.tremula* x *P.alba* et *P.trichocarpa*, il n'en est pas de même pour les régions intergéniques. Dans ce chapitre, les études d'expression ont porté sur les deux clones.

1- Etude d'expression

1.1- MYB090

L'expression de MYB090 a d'abord été suivie par RT-PCR sur différents tissus prélevés sur un arbre de 4 ans (717-1B4) (figI.2). Des fleurs mâles et femelles ont par ailleurs été prélevées sur des individus *P. nigra* arrivés à floraison. *MYB090* est fortement exprimé dans le xylème secondaire du tronc et des racines de l'arbre de 4 ans, surtout dans le xylème en différenciation, mais également dans le xylème mature de la tige. On détecte une expression de *MYB090* dans les feuilles, surtout dans les pétioles. On ne détecte aucune expression dans le phloème secondaire, que ce soit dans le tronc ou les racines, et un très léger signal dans les apex, les zones cambiales de tige et de racine, et les fleurs. Dans de jeunes tiges de 717-1B4 inclinées, on détecte *MYB090* à la fois dans le bois opposé et le bois de tension, avec une possible surexpression dans le bois de tension. Cette surexpression a été confirmée par Q-PCR sur le clone 101-74 : l'expression de *MYB090* est deux fois plus élevée dans le bois de tension par rapport au bois opposé (figI.3).

Nous avons également exploité les données disponibles en ligne, notamment sur le Poplar eFPBrowser (<http://bar.utoronto.ca/efppop/cgi-bin/efpWeb.cgi>) qui permet de disposer d'une représentation graphique de données d'expression obtenues sur puce Affymetrix à partir d'ARN préparés à partir de différents tissus / traitements sur différents individus de *P. balsamifera* (Wilkins *et al.* 2009). Sur ces échantillons, on retrouve une expression forte dans le xylème secondaire des tiges, mais également une expression dans les chatons mâles. L'expression la plus forte est détectée dans les chatons femelles (figI.4).

Poplar eFP Browser at bar.utoronto.ca

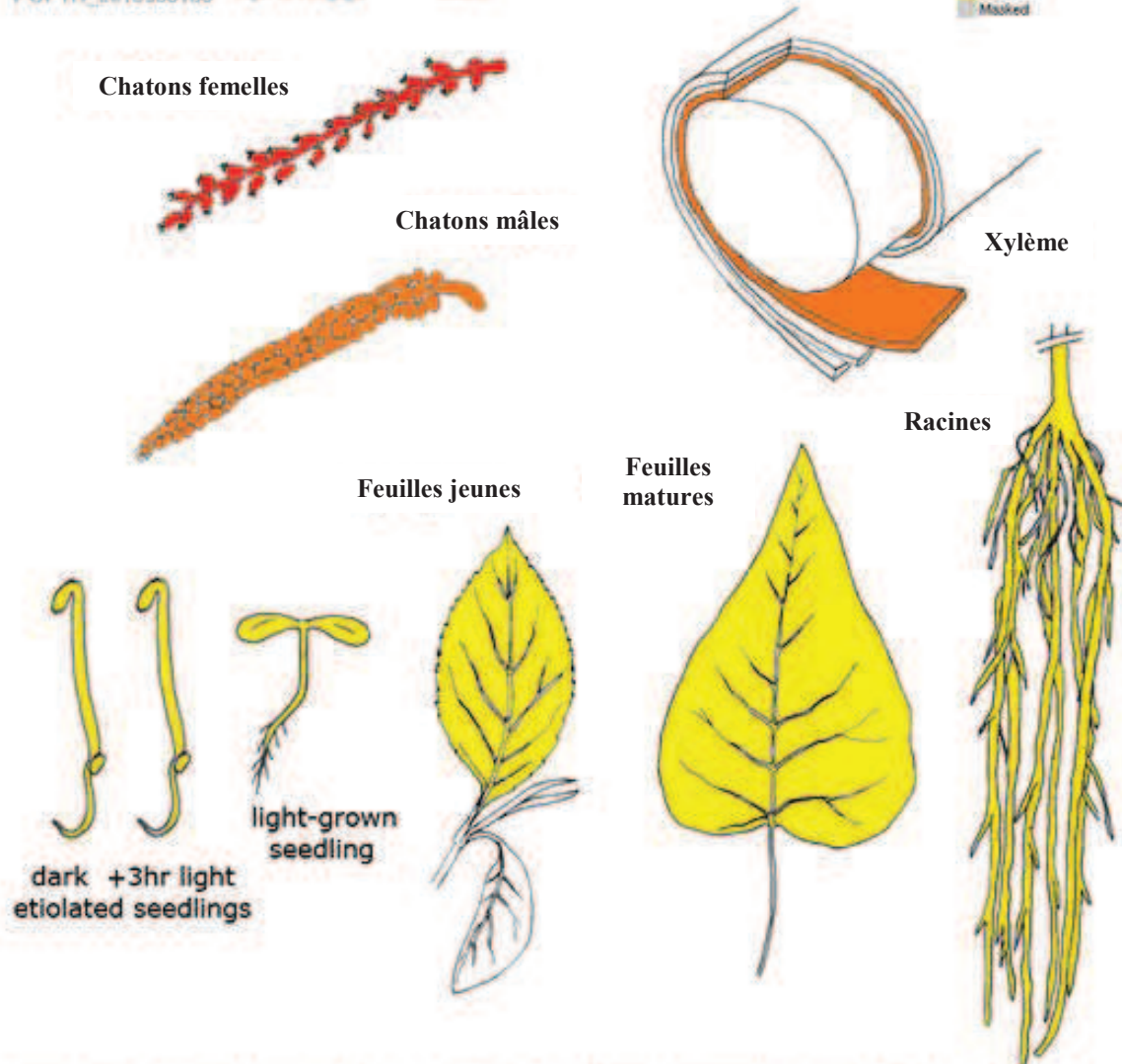
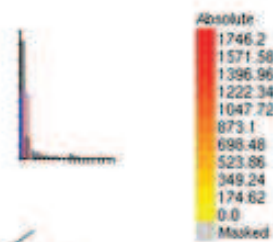
Wilkins et al., 2008. Plant Physiol. 149:981-993

Poplar eFP Browser Developmental Series. Data from the Campbell Laboratory. Affymetrix expression data normalized by the GCOS method, with a TGT value of 500. Duplicate or triplicate samples were analyzed from greenhouse-grown or field-grown material (in the case of the catkins). The seedlings were grown on moist filter paper. All material was grown under a diurnal cycle of 12h light/dark and sampled at midday, except for the xylem samples, which were sampled at midnight.

POPTR_0015s05130

Ptp.116.1.S1_s_at

afv652



Images drawn by Josephine McKeever and Nicholas Provart. Poplar eFP Browser implemented by Justin Foong and Hardeep Nahal.

Figure I.4 : Représentation graphique de l'expression de *MYB090* au sein de différents tissus et/ou traitements chez différents *P. trichocarpa*. L'expression est mesurée par hybridation sur puces Affymetrix Peuplier (eFP browser, <http://bar.utoronto.ca/efppop/cgi-bin/efpWeb.cgi>, Wilkins et al. 2008). La gamme de couleur va du jaune (faible accumulation de transcrits) au rouge (forte accumulation de transcrits). *MYB090* correspond au modèle de gène Potri.015G033600 (version génome v3.0) ou POPTR_0015s05130 (version génome v2.0). Il est représenté sur la puce Affymetrix Peuplier par la sonde Ptp.116.1.S1_s_at.

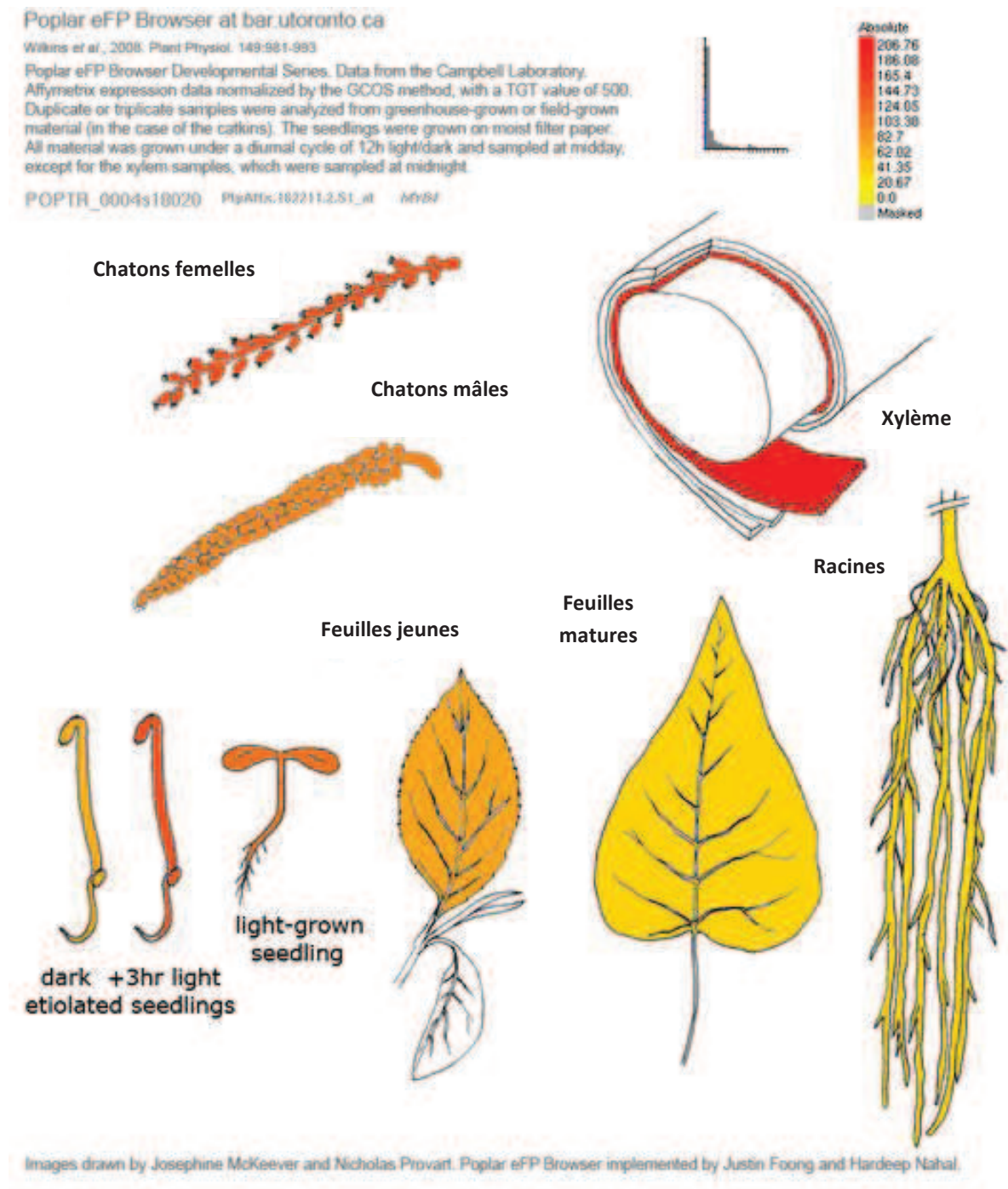


Figure I.5 : Représentation graphique de l'expression de *MYB221* au sein de différents tissus et/ou traitements chez différents *P. trichocarpa*.

L'expression est mesurée par hybridation sur puces Affymetrix Peuplier (eFP browser, <http://bar.utoronto.ca/efppop/cgi-bin/efpWeb.cgi>, Wilkins et al. 2008). La gamme de couleur va du jaune (faible accumulation de transcrits) au rouge (forte accumulation de transcrits). MYB221 correspond au modèle de gène Potri.004G174400 (version génome v3.0) ou POPTR_0004s18020 (version génome v2.0). Il est représenté sur la puce Affymetrix Peuplier par la sonde PtpAffx.162211.2.S1_at.

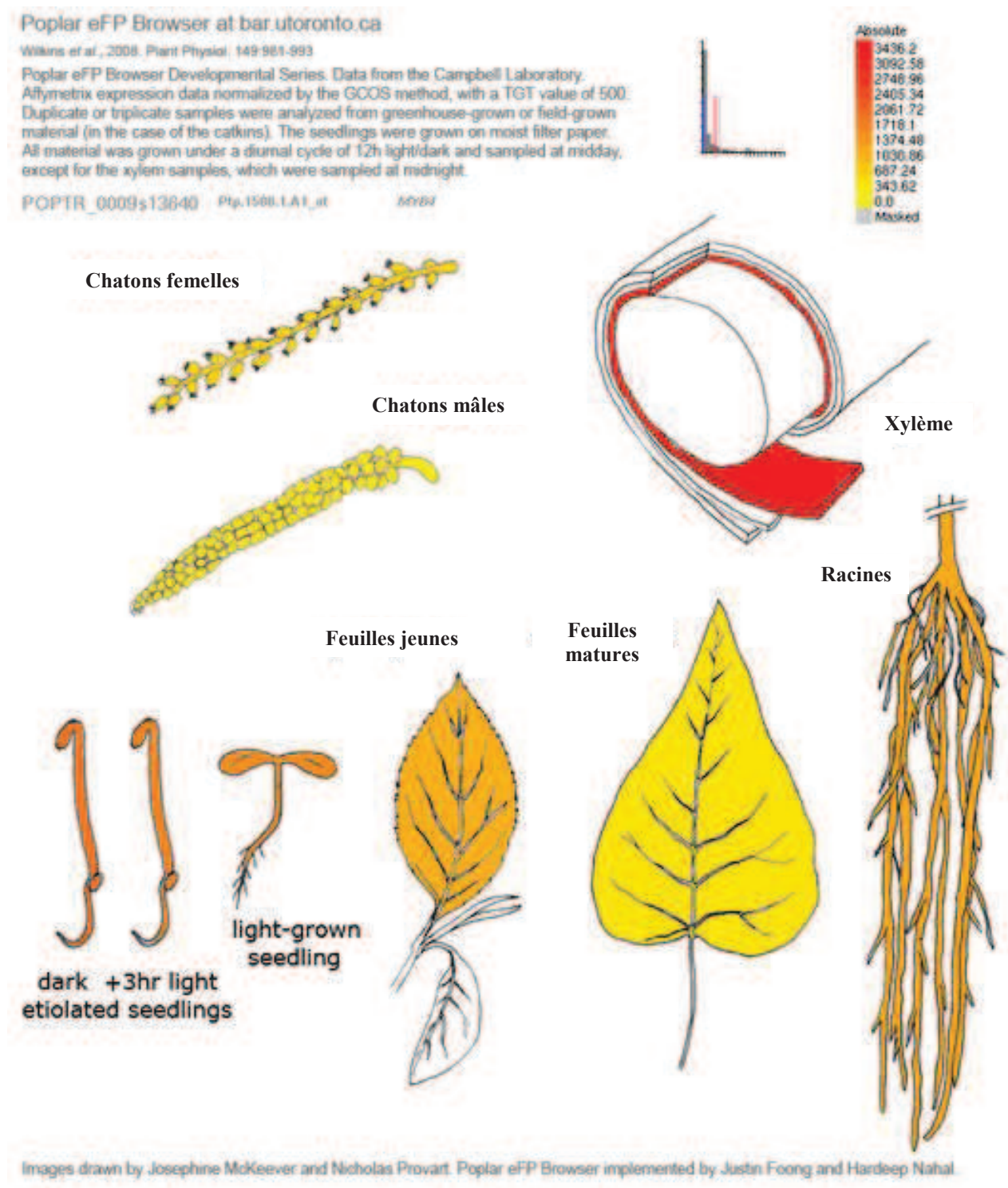


Figure I.6 : Représentation graphique de l'expression de *MYB156* au sein de différents tissus et/ou traitements chez différents *P. trichocarpa*. L'expression est mesurée par hybridation sur puces Affymetrix Peuplier (eFP browser, <http://bar.utoronto.ca/efppop/cgi-bin/efpWeb.cgi>, Wilkins et al. 2008). La gamme de couleur va du jaune (faible accumulation de transcrits) au rouge (forte accumulation de transcrits). *MYB156* correspond au modèle de gène Potri.009G134000 (version génome v3.0) ou POPTR_0009s13640 (version génome v2.0). Il est représenté sur la puce Affymetrix Peuplier par la sonde Ptp.1500.1.A1_at.

1.2- MYB156 et MYB221

MYB156 et MYB221 ont été choisis parce que ce sont des orthologues d'EgMYB1, un facteur MYB d'eucalyptus, répresseur de la voie de biosynthèse des lignines, et dont certaines cibles ont été démontrées expérimentalement (Legay *et al.* 2007). Notamment, EgMYB1 a pour cible les promoteurs des gènes *CAD* et *CCR* (Legay *et al.* 2007). Nous faisons l'hypothèse que MYB156 et/ou MYB221 ont la même fonction chez le peuplier et les mêmes cibles. Cela nous sera utile pour la mise au point et la validation des expériences de ChIP-SEQ (voir Chapitre 2). *MYB156* et *MYB221* ont globalement le même profil d'expression chez *P. trichocarpa* (Poplar eFP browser), avec notamment une expression maximale dans le xylème secondaire de la tige (figI.5 et I.6). Il existe toutefois quelques différences notables : l'expression de *MYB221* est relativement fortement détectée dans les fleurs, femelles et mâles, alors que celle de *MYB156* n'est pas détectée. En valeur absolue, *MYB156* est exprimé beaucoup plus fortement que *MYB221*.

Nous avons étudié le profil d'expression de *MYB221* et de *MYB156* dans les cellules de xylème de bois de tension et de bois normal.

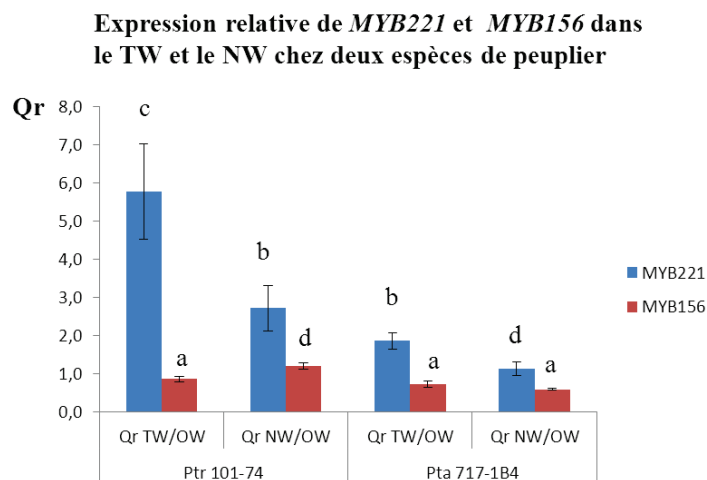


Figure I.7 : Expression relative de *MYB221* et *MYB156* dans le bois de tension (TW) ou le bois normal (NW) chez deux espèces différentes de peuplier. L'expression quantitative relative (Qr) de MYB221 (rectangles bleus) et MYB156 (rectangles rouges) dans le xylème tendu (TW) et normal (NW) a été calculée par rapport au bois opposé (OW). Le Qr a été calculé chez deux espèces de peuplier: *P. trichocarpa* clone 101-74 (Ptr 101-74) et *P. tremula x P. alba* clone INRA 717-1B4 (Pta 717-1B4). Les barres d'erreur représentent l'erreur type de la moyenne de 4 répétitions biologiques pour Ptr 101-74 et 3 répétitions biologiques pour Pta 717-1B4. Les lettres différentes indiquent une différence significative entre les valeurs de Qr (t-test $p < 0.05$).

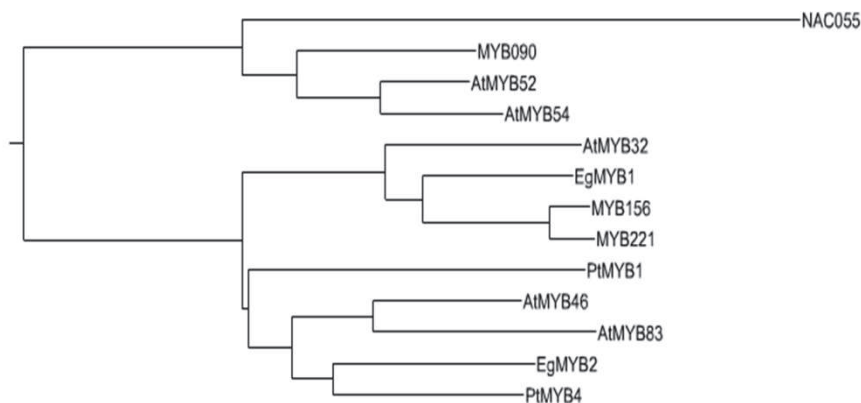


Figure I.8 : Analyse phylogénétique de 13 séquences nucléotidiques codant pour des facteurs de transcription impliqués dans la formation de la paroi secondaire.

Le clustering a été réalisé selon la méthode de Neighbour-Joining. Les distances ont été corrigées et les gaps ont été exclus. Le calcul a été réalisé sur ClustalW2 et l'arbre a été élaboré par ITOL (Letunic & Bork, 2007).

CLUSTAL 2.1 multiple sequence alignment

```

Potri.015G033600.1  --MCTRGHWRPAEDEKLKELVEKYGPHNWNIAAEKLHGRSGKSCRLRWFNQLDPRINRSP  58
PtaMyb090          --MCTRGHWRPAEDEKLKELVEKYGPHNWNIAAEKLHGRSGKSCRLRWFNQLDPRINRSP  58
AtMYB52            -MCSRGRHWRPAEDEKLRELVEQFGPHNWNIAAQKLSGRSGKSCRLRWFNQLDPRINRNP  59
AtMYB54            MIMCSRGRHWRPAEDEKLKDLVEQYGPHNWNIAALKLPGRSGKSCRLRWFNQLDPRINRNP  60
                    **:*****::***:***** ** *****

```

```

Potri.015G033600.1  FTEEEERLLASHRIHGNRWAVIARLFPGRITNAVKNHWHVIMARRYRERSR--LHAKR- 115
PtaMyb090          FTEEEERLLASHRIHGNRWAVIARLFPGRITNAVKNHWHVIMARRYRERSR--LHAKR- 115
AtMYB52            FTEEEERLLASHRIHGNRWSVIARFFPGRITNAVKNHWHVIMARRGRERSK--LRPRG- 116
AtMYB54            FTEEEERLLAAHRIHGNRWSIARLFPGRITNAVKNHWHVIMARRTRQTSKPRLLPSTT 120
                    *****:*****:***:*****

```

```

Potri.015G033600.1  -TAQALVNEQKFSSQDMQINCETRSFSSFVKKYCEKFG-QYPLITHS-YLPAFWKEFYN 172
PtaMyb090          -TAQALVNEQNFFSSQDMQINCETRSFSSFVKKYCEKFG-QYPLITHS-YLPAFWKEFYN 172
AtMYB52            -----LGHDGIVAATGMIGNYKDCDKERRLATTTAINFPYQFSHINHFQVLKEFLTGKIG 171
AtMYB54            SSSSLMASEQIMMSSGGYNHNYSSDDRKKIFPADFINFPYKFSHINHLHFLKEFFTGKIA 180
                    : : . : . : : * :. *. * * .

```

```

Potri.015G033600.1  -DQLSNCK-----DQNRPIEFYDFLQVNTESNKSEVIDNARREDEEVDQREVILEHQSK 225
PtaMyb090          GDDPSNCE-----DQNRSIEFYDFLQVNTDSNKSEVIDNARREDEEVDQQEVILEHQSK 226
AtMYB52            FRNSTPIQEGAIQTKRPMEFYNFLQVNTDSKIHELIDNSRKDEEEDVDQNNRIPNEN- 230
AtMYB54            LNHKAN-----QSKKPMEFYNFLQVNTDSNKSEIID---QDSGQSKRSDSSTKHES- 228
                    . :. : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Potri.015G033600.1  AGVPFIDFFSA----- 236
PtaMyb090          AGVPFLIFSQ----- 237
AtMYB52            -CVPFFDFLSVGNSSASQGLC 249
AtMYB54            -HVPFFDFLSVGNSSAS---- 243
                    ***: * .

```

Figure I.9 : Alignement de la séquence protéique de MYB90 (*P. tremula* x *P. alba*) avec ses orthologues chez *P. trichocarpa* et chez *Arabidopsis thaliana*.

En rouge et bleu sont indiqués les deux domaines Pfam PF00249 MYB-like DNA binding domain

La figure I.7 montre le niveau d'expression relative de *MYB221* chez 101-74 (*Populus trichocarpa*) et chez 717-1B4 (*Populus tremula x alba*) : il est plus élevé dans le xylème tendu que dans le xylème normal, presque 6 fois plus élevé chez 101-74 et seulement 2 fois plus fort chez 717-1B4. Par comparaison, le niveau d'expression de *MYB156* ne varie pratiquement pas entre xylème de bois tendu, de bois opposé ou de bois normal (figI.7) quel que soit le clone considéré. En valeur absolue, contrairement à ce qui a été observé chez *P.trichocarpa*, c'est *MYB221* qui est plus fortement exprimé dans le jeune xylème, par rapport à *MYB156* : on a des valeurs de Ct de 24 pour *MYB156*, 22,5 pour *MYB221* pour un Ct de 19 pour le gène de référence Ubiquitine.

2- Clonage et analyse des séquences protéiques

Les séquences codantes de *MYB090*, *MYB156* et *MYB221* ont été clonées par PCR à partir d'une matrice d'ADNc préparée à partir d'ARN extraits de jeune xylème de bois de tension de 717-1B4 (*Populus tremula x P. alba*). Les amorces ont été dessinées d'après les séquences des transcrits identifiés chez Nisqually (*P. trichocarpa*), respectivement Potri.015G033600.1, Potri009G134000.1 et Potri004G174400. Les produits PCR ainsi amplifiés ont été clonés dans le vecteur pCR2.1 et vérifiés par séquençage. Les séquences obtenues pour 717-1B4 ont été comparées aux séquences de Nisqually. Elles sont notées PtaMYBXXX. Les motifs caractéristiques des FT de type MYB ont été recherchés.

2.1- MYB090

MYB090 est un orthologue d'*AtMYB52* et *ATMYB54* (figI.8). Les trois gènes dérivent d'un même ancêtre, mais chez *Arabidopsis*, il y a eu duplication de gènes. Nous avons réalisé un alignement de séquences multiples en utilisant Clustal W (Thompson *et al.* 1994) à partir de la séquence PtaMYB090, son équivalent chez *P. trichocarpa*, *AtMYB52* et *AtMYB54* (figI.9). Les deux séquences de peuplier sont très proches, avec 98% d'identité au niveau protéique. La région N-terminale est très bien conservée entre les 4 séquences et contient les deux domaines « MYB-like DNA binding domain » R2 et R3 (domaine PFAM PF00249, <http://pfam.sanger.ac.uk/>). *MYB090* appartient donc à la famille des FT MYB de type R2-R3. Dans la partie C-terminale, une zone de 16-20 acides aminés est également bien

Accession	Protein	Sequence	Score
PtaMYB156	MGRSPCCCKAHTNKGAWTKEEDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY	60	
Potri.009G134000.1	MGRSPCCCKAHTNKGAWTKEEDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY	60	
PtaMYB221	MGRSPCCCKAHTNKGAWTKEEDRLIAYIRTHGEGCWRSLPKAAGLLRCGKSCRLRWINY	60	
Potri.004G174400.1	MGRSPCCCKAHTNKGAWTKEEDRLIAYIRTHGEGCWRSLPKAAGLLRCGKSCRLRWINY	60	
EgMYB1	MGRSPCCCKAHTNKGAWTKEEDKLIAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY	60	
AtMYB4	MGRSPCCCKAHTNKGAWTKEEDRLVAYIRAHGEGCWRSLPKAAGLLRCGKSCRLRWINY	60	
*****:::***::*****			
PtaMYB156	LRPDLKRGNFTEAEDELIIKLHSLLGKNSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRG	120	
Potri.009G134000.1	LRPDLKRGNFTEAEDELIIKLHSLLGKNSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRG	120	
PtaMYB221	LRPDLKRGNFTEAEDELIIKLHSLLGKNSLIAGRLPGRTDNEIRNYWNTHIRRKLLNRG	120	
Potri.004G174400.1	LRPDLKRGNFTEAEDELIIKLHSLLGKNSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRG	120	
EgMYB1	LRPDLKRGNFTEAEDEIIKLHSLLGKNSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRG	120	
AtMYB4	LRPDLKRGNFTEAEDELIIKLHSLLGKNSLIAGRLPGRTDNEIKNYWNTHIRRKLLNRG	120	
***** **::*****::*****::***			
PtaMYB156	IDPATHRPLNEPAV-QEATTT-----ISFTTTTTSVLEESLSGSIKEENKEKI	168	
Potri.009G134000.1	IDPATHRPLNEPAV-QEATTT-----ISFTTTTTSVLEESLSGSIKEENKEKI	168	
PtaMYB221	IDPATHRPLNEPA--QEASTT-----ISFSTTTS--VKDESLSVKEESNKEKI	165	
Potri.004G174400.1	IDPATHRPLNEPA--QEASTT-----ISFSTTTS--VKDESLSVKEESNKEKI	165	
EgMYB1	IDPATHRLINEPAQDHHDEPT-----ISFAANSK-----EIKEMKN--NAELN	161	
AtMYB4	IDFTSHRPIQESSAQDSKPTQLFVTSNTIINISFTSAPK----VETFHESISFGPKSEK	176	
::** ::*: :: . * *: .. : . . .			
PtaMYB156	ISATAFVCKEEKTKVQERCSDLNLELGISLPSQNQPDHQPFFKTGGSRSLCFACSLGLQ	227	
Potri.009G134000.1	ISATAFVCKEEKTKVQERCSDLNLELGISLPSQNQPDHQPFFKTGGSRSLCFACSLGLQ	227	
PtaMYB221	ISAAAFICKEEKNFVQERCPLDNLELRISLPCQNQPDHQAFTKGSTSLCFACSLGLQ	222	
Potri.004G174400.1	ISAAAFICKEEKTFVQERCPLDNLELRISLPCQNQPDHQAFTKGSTSLCFACSLGLQ	224	
EgMYB1	FMCNLEESADVASSARERCPLDNLELGISPPSHQLHQPEPLRFTGRKSDLCXECNGLK	221	
AtMYB4	ISMLTFKEEKDECPVQEKFPDLNLELRISLP----DDVDRLQGHGKSTTPRCFKCSLGM	232	
: . .::: .***** ** * : : * *.***			
PtaMYB156	NSKDCSCSVI-VSTVGSSG-STSIKTGYDFLGMK----SGVLDYRSLEMK	271	
Potri.009G134000.1	NSKDCSCSVI-VSTVGSSG-STSIKTGYDFLGMK----SGVLDYRSLEMK	271	
PtaMYB221	NSKDCSCSVI-VGTIGSSS-SAGSKTGYDFLGMK----SGVLDYRGLEMK	268	
Potri.004G174400.1	NSKDCSCSVI-VGTIGSSS-SAGSKTGYDFLGMK----SGVLDYRGLEMK	268	
EgMYB1	NSQNCRCS---VGVISS-----ETSVGYDFLGK----ASVLDYRS---255		
AtMYB4	NGMECRCSGRMRCDDVVGSSSGSDMSNGDFLGKAKETISLLGRFSLEMK	282	
* . * * * :*****: .:::.*.			

En rouge et bleu sont indiqués les deux domaines Pfam PF00249 MYB-like DNA binding domain. En mauve est indiqué le motif EAR de répression de la transcription. Les flèches vertes indiquent les quatre cystéines impliquées dans une structure en doigt de zinc.

conservée entre les 4 séquences : les 2 séquences de peuplier sont plus proches d'AtMYB52 que d'AtMYB54. Une analyse de la séquence de PtaMYB090 sur PSORT montre que la séquence ne contient pas de signal caractéristique d'adressage au noyau (<http://psort.hgc.jp/>).

2.2- MYB156 et MYB221

Les séquences protéiques de MYB156 et MYB221 ont été alignées avec leurs orthologues chez *P. trichocarpa*, ainsi que leurs orthologues chez Eucalyptus et Arabidopsis, EgMYB1 et AtMYB4, respectivement (figI.10). Les quatre séquences de peuplier sont très proches : PtaMYB156 et PtaMYB221 ont 86% d'identité entre elles et chacune 99% avec leur orthologue chez *P. trichocarpa*. La région N-terminale est très bien conservée pour toutes ces séquences, notamment sur les 125 premiers acides aminés. Cette région contient les deux domaines «MYB-like DNA binding domain» (domaine PFAM PF00249, <http://pfam.sanger.ac.uk/>), caractéristiques des MYB de type R2-R3. Dans le motif R3, on retrouve la séquence [DE]Lx₂[RK]x₃Lx₆Lx₃R nécessaire à l'interaction avec un facteur bHLH (Zimmermann *et al.* 2004). En région C-terminale, on retrouve pour toutes les protéines un motif de type EAR (LxLxLx ou DLNxxP), impliqué dans la répression de la transcription (Kagale et Rozwadowski, 2011). On trouve également pour toutes les protéines un motif Cx₂Cx₉CxC, motif à doigt de zinc.

3- Production d'anticorps et étude protéique

Des anticorps polyclonaux pour MYB090 et MYB221 ont été produits à partir de protéines recombinantes. MYB221 a été choisi par rapport à MYB156 car il est surexprimé dans le xylème de bois de tension. Les séquences codantes complètes de MYB090 et MYB221 ont été amplifiées par PCR avec des amorces de type Gateway®, et ont été introduites par recombinaison dans le vecteur d'entrée pDONR207 (Invitrogen™), puis transférées par recombinaison dans différents vecteurs spécifiques permettant la production de protéines recombinantes avec une étiquette 6His dans *E. coli*. MYB90 et MYB221 ont été clonés respectivement dans pETG-10K (Alzari *et al.* 2006) et pDEST17 (Invitrogen™). Les protéines ont d'abord été solubilisées dans une solution d'urée 4M puis purifiées par affinité

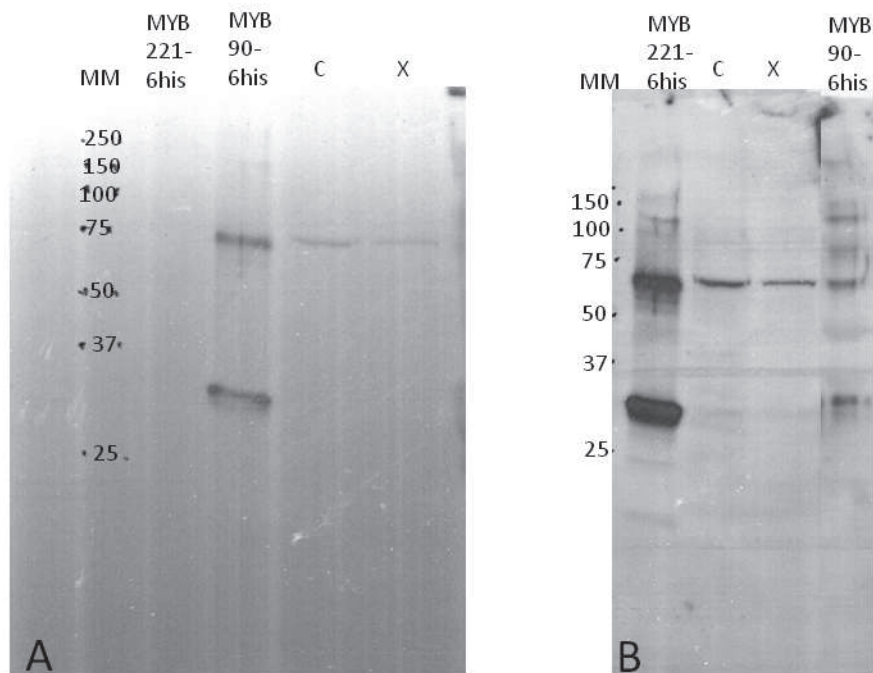


Figure I.11 : Détection par Western blot de MYB090 et MYB221 dans des extraits de protéines nucléaires de cambium (C) ou de jeune xylème (X), avec des anticorps anti-MYB090 (A) et anti-MYB221 (B)

Extrait de protéines nucléaires de cambium (C) et xylème (X) : 55µg/puits ; Myb90-6his et Myb221-6his recombinants : 50 ng/puits. Anticorps primaires polyclonaux, purifiés par affinité sur antigène anti-Myb90(A) et anti-Myb221 (B): dilution 1:2000 ; Anticorps secondaire anti-IgG lapin-HRP: dilution 1:3000 et réactifs du kit Amersham ECL Western blotting. Durée exposition : 15 minutes

sur une colonne de cobalt, éluées en imidazole. Les éluats ont été dialysés pour éliminer l'urée.

Les anticorps polyclonaux anti-MYB90 et anti-MYB221 ont été produits chez le lapin par la société Agro-Bio (La Ferté Saint Aubin). Nous leur avons également confié la purification des anticorps par affinité sur antigène. Dans des échantillons de protéines nucléaires, l'anticorps anti-MYB090 détecte 2 bandes, 1 bande de très faible intensité autour de la taille attendue de 30 kDa, et une bande de plus forte intensité, à environ 70 kDa. L'anticorps anti-MYB221 donne le même type de détection (figI.11). L'anticorps anti-MYB090 ne reconnaît pas la protéine recombinante MYB221, alors que l'anticorps anti-MYB221 détecte légèrement la protéine recombinante MYB090 (figI.11).

4- Localisation subcellulaire des FT

Les séquences codantes de MYB221, MYB156 et MYB090 ont été clonées dans le vecteur pSCA ::35S ::YFP permettant la production de FT fusionné à la protéine fluorescente Yellow Fluorescent Protein (YFP), afin de tester leur localisation sub-cellulaire (Guirimand *et al.* 2010). Ces vecteurs ainsi qu'une construction portant un marqueur de noyau ont été introduits dans des cellules de *Catharanthus roseus* (pervenue de Madagascar) par biolistique ; leur expression a été suivie de façon transitoire. Les FT produits ont pu être localisés au niveau de ces cellules grâce à la fluorescence de YFP et CFP. La figure I.12 montre une superposition parfaite entre la fluorescence de FT-YFP et la fluorescence du marqueur de noyau (CFP) indiquant clairement que les 3 FTs sont localisés exclusivement dans le noyau car aucune fluorescence n'a été détectée dans le cytoplasme.

5- Conclusions

Les 3 facteurs MYB choisis appartiennent tous les 3 à la classe des MYB R2 R3 car on détecte la présence de 2 motifs MYB DNA binding domain. Il y a 192 gènes annotés comme tels chez *Populus* (126 chez *Arabidopsis*, Wilkins *et al.* 2009). Ces deux domaines sont des répétitions imparfaites de 50 acides aminés environ formant une structure en hélice/boucle/hélice impliquée dans la fixation du FT à l'ADN. Les protéines ne présentent pas de signal d'adressage au noyau caractéristique mais leur localisation nucléaire a été

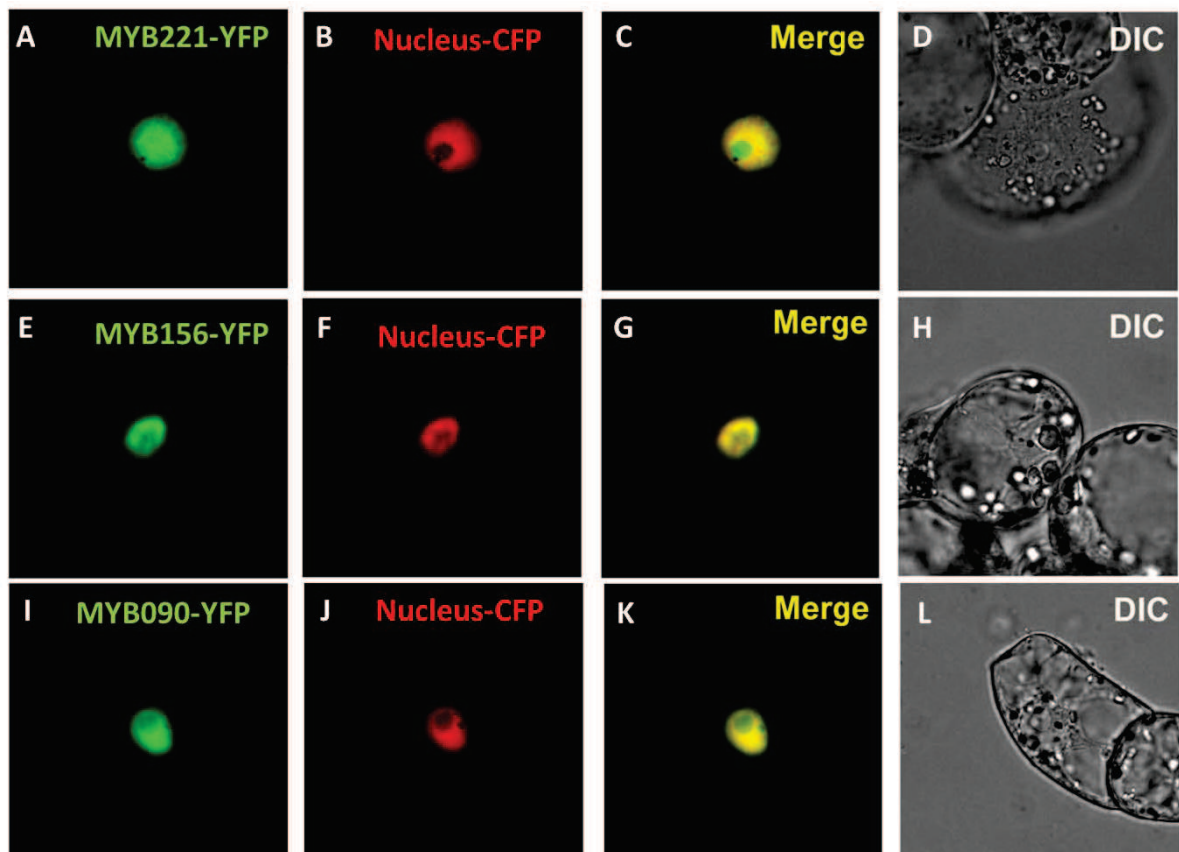


Figure I.12: Localisation subcellulaire de MYB221-YFP, MYB156-YFP et MYB090-YFP dans des cellules de *Catharanthus roseus*.

Les cellules ont été co-transformées à l'aide des plasmides exprimant MYB221-YFP (A), MYB156-YFP (E) ou MYB090-YFP (I) d'une part, et le marqueur nucléaire CFP d'autre part (B, F, J). La colocalisation des deux signaux de fluorescence est démontrée au niveau des images superposées (C, G, K). La morphologie des cellules est observée par microscopie DIC (differential interference contrast) (D, H, L).

clairement démontrée en système cellulaire hétérologue (fig.I.12). Les séquences clonées sont suffisantes pour un adressage correct des protéines produites au noyau

MYB221 et *MYB156* sont des gènes codant pour des FT qui sont des orthologues d'*EgMYB1* et *AtMYB4* (Ma *et al.* 2011), connus pour être des répresseurs de la transcription de gènes de biosynthèse des lignines chez *Eucalyptus* et *Arabidopsis* (Legay *et al.* 2010, Jin *et al.* 2000). Ces deux FT appartiennent au sous-groupe 4 de la sous-famille MYB-R2R3. Les FT appartenant à ce sous-groupe sont des répresseurs de la voie de biosynthèse des phenylpropanoïdes. *MYB221* et *MYB156* ont bien la signature caractéristique de répression de la transcription (motif EAR, fig.I.10). Les séquences protéiques de *MYB221* et *MYB156* sont très proches (86% d'identité). Les deux gènes résultent d'une duplication récente, qui s'est produite de façon ultérieure à la divergence des espèces *Arabidopsis/P. trichocarpa* : y-a-t-il eu fonctionnalisation des deux gènes depuis leur duplication ou sont-ils fonctionnellement redondants? Au niveau expressionnel, on note des différences entre les deux gènes, avec notamment une sur-expression de *MYB221* dans le bois de tension, qui n'est pas observée pour *MYB156*. C'est pourquoi *MYB221* a été retenu pour la production d'anticorps et la mise en œuvre de la ChIP (Chapitre 2). L'hypothèse de travail est que *MYB221* joue un rôle dans la baisse des teneurs en lignines dans le bois de tension. Pour les études *in planta* (Chapitre 3), nous avons cependant choisi de travailler en parallèle sur les deux FT pour mettre en évidence leur fonction éventuellement différente dans la physiologie de la plante.

MYB090 appartient au même groupe phylogénétique qu'*AtMYB52* et *AtMYB54*. Zhong *et al.* (2008) ont proposé qu'*AtMYB52* soit un activateur transcriptionnel de la voie de biosynthèse des lignines car des plants sur-exprimant une version dominante répresseur d'*AtMYB52* présentent une réduction notable de l'épaisseur des parois secondaires des fibres interfasciculaires dans les tiges d'*Arabidopsis*. Chez *P. tremula x P. tremuloides*, il a été proposé par contre que *PttMYB21a*, orthologue de *MYB090*, soit un répresseur du gène de CCoAOMT. Nous ne pouvons à ce stade faire d'hypothèse claire sur la fonction de *MYB090*, les travaux du chapitre 3 devraient nous permettre de répondre.

Chapitre 2

Identification des cibles ADN des facteurs de transcription à l'échelle du génome

Pour déterminer les cibles des FT MYB090 et MYB221 à l'échelle du génome, nous avons utilisé la technique d'Immunoprécipitation de la Chromatine couplée à un séquençage haut débit des fragments immunoprécipités (ChIP-SEQ). Dans ce chapitre, nous introduirons cette technique d'une façon générale puis nous détaillerons sa mise en œuvre sur notre matériel d'étude, le jeune xylème de peuplier, avant de présenter nos résultats.

1- Principes de l'immunoprécipitation de la chromatine

L'immunoprécipitation de la chromatine (ChIP) est une technique largement utilisée dans le domaine animal pour déterminer, *in vivo*, la localisation des sites de fixation à l'ADN de différents facteurs de transcription mais aussi des histones et d'autres protéines. Cette technique aide à la détection des interactions protéines-ADN dans des cellules vivantes car les protéines sont "capturées" au niveau de leurs sites effectifs de fixation aux acides nucléiques (Das *et al.* 2004). Cette technique a été utilisée pour la première fois chez la drosophile (Gilmour et Lis 1985). Chez les plantes, elle a été employée pour la première fois chez *Arabidopsis* en 1999 (Ascenzi et Gantt 1999). Cette technique a d'abord été mise en œuvre sur des cultures cellulaires, donc des cellules avec une paroi peu épaisse, ce qui est très différent du tissu sur lequel on travaille.

La ChIP consiste généralement en une première étape permettant la liaison des protéines de manière covalente sur la chromatine, à l'aide d'un agent de pontage qui peut être chimique comme le formaldéhyde, le chromate de potassium, le cisplatine ou de la lumière UV et des lasers. Cette étape est suivie par une sonication ou un traitement avec une nucléase pour générer des petits fragments d'ADN. L'ADN correspondant aux sites de fixation des protéines étudiées est protégé de la sonication par le pontage. L'immunoprécipitation est ensuite réalisée à l'aide d'anticorps spécifiques des protéines d'intérêt, FT ou histones. L'ADN est par la suite libéré des protéines immunoprécipitées et analysé à l'aide de techniques courantes de biologie moléculaire, telles que la PCR, la qPCR quand on connaît des cibles ADN potentielles : on va alors mesurer des enrichissements spécifiques en cibles ADN connues, pour lesquelles on a dessiné des amorces (pour revue voir Das *et al.* 2004).

Il existe deux types d'essais de ChIP : la X-ChIP et la N-ChIP. La méthode X-ChIP utilise la chromatine pontée fragmentée par sonication alors que la N-ChIP utilise de la chromatine native non pontée et digérée par une nucléase. Dans les essais N-ChIP, l'immunoprécipitation est très efficace et aboutit à des quantités importantes d'ADN immunoprécipité. En effet, la liaison des anticorps aux protéines non pontées est généralement plus forte qu'aux protéines pontées. C'est ainsi que la spécificité de l'immunoprécipitation de la N-ChIP est plus grande que celle de la X-ChIP. La N-ChIP convient surtout pour l'étude de protéines fermement liées à la chromatine comme les histones. Ainsi, elle est déconseillée pour les protéines dont les liaisons avec l'ADN sont plus labiles, et donc susceptibles d'être modifiées durant l'expérience. Dans ce dernier cas, la X-ChIP est plus adaptée car le pontage fixe de façon covalente les interactions ADN/protéines et protéines/protéines. Toutefois, en cas d'excès de fixation, la fragmentation de l'ADN peut être difficile (pour revue voir Das *et al.* 2004).

La ChIP peut être aussi classée selon les techniques utilisées pour l'analyse des séquences ADN obtenues. Dans le cas où ces dernières sont analysées par PCR on parle alors de ChIP-PCR. De même, par puces à ADN : ChIP-chip, par clonage (ChIP-Cloning)... (Das *et al.* 2004).

2- Détermination des cibles de MYB221et MYB090 à l'échelle du génome

Notre objectif est de déterminer *in vivo* les cibles de MYB090 et MYB221 dans le cadre de la formation du bois de tension pour les raison énumérées dans le chapitre 1 de ce manuscrit. Il s'agit de deux facteurs de transcription, et donc de protéines qui ne sont pas liées de façon permanente à la chromatine comme les histones. Ainsi la X-ChIP convient plus pour notre étude. De plus, le développement récent des technologies de séquençage haut-débit a permis d'utiliser cet outil à la suite de la ChIP. On parle alors de ChIP-SEQ, technique qui permet l'identification des sites de fixation à l'ADN à l'échelle du génome entier par le séquençage direct de l'ADN récupéré par immunoprécipitation (Barski et Zhao 2009, Soler *et al.* 2011). C'est ainsi qu'on a choisi la ChIP-SEQ pour répondre à notre objectif.

2.1- Mise en œuvre de la technique sur notre matériel biologique

Cette partie ne constitue pas une répétition du Matériel et Méthodes mais une synthèse de la procédure suivie dans notre étude avec notification des éventuels points-clés, explications et remarques importantes sans entrer dans les détails techniques (produits utilisés et concentrations, étapes détaillées ...).

2.1.1- Pontage de la chromatine *in vivo*

Nous avons prélevé des moitiés de fragments de tige écorcées correspondant au bois de tension. Pour les ponter, nous avons utilisé le formaldéhyde car il est le plus couramment utilisé dans la littérature comme agent de pontage entre protéines et ADN. C'est une petite molécule, qui crée des liaisons covalentes entre groupements amines présents sur certains nucléotides et sur certains acides aminés comme la lysine, sur des distances très faibles de l'ordre de 0,2 nm. Elle a ainsi la capacité de fixer des molécules en étroite association les unes avec les autres. Ces liaisons ont l'avantage d'être réversibles. Cet agent chimique engendre le pontage des protéines avec l'ADN, mais également des protéines avec l'ARN et des protéines avec d'autres protéines.

Des essais préliminaires ont permis à mon équipe d'accueil de déterminer les conditions optimales de pontage de nos échantillons. Le premier critère important est de faire le pontage par infiltration sous vide de la solution de formaldéhyde, ce qui permet de franchir plus facilement la barrière physique que constitue la paroi cellulaire. Le temps optimal de pontage de nos échantillons a été établi à 20 min, suivi d'une infiltration sous-vide d'une solution de glycine pendant 5 min pour arrêter le pontage et éviter une sur-fixation des tissus.

2.1.2- Sonication de la chromatine pontée

Les noyaux sont purifiés à partir de jeune xylème tendu prélevé par grattage sur les tiges après pontage. Ensuite, la chromatine est extraite à l'aide d'un tampon de lyse contenant du SDS (important pour la sonication par ultrasons). Mon équipe d'accueil a réalisé des essais pour choisir les conditions optimales de sonication. Cette étape est réalisée donc pendant 30 min avec des pulses de 30s de sonication espacés de 30 s.

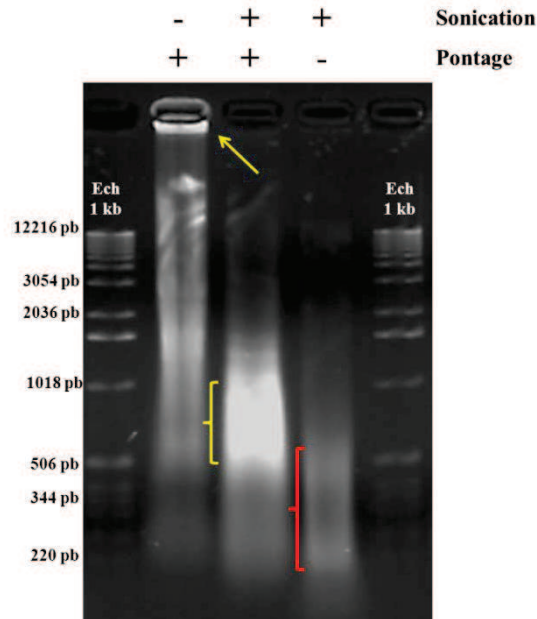


Figure II.1 : Vérification de la qualité de sonication sur gel d'agarose.

Dépôt sur gel des aliquotes de : chromatine pontée non soniquée (-/+), chromatine pontée soniquée (+/+) et chromatine soniquée dépontée (+/-). L'échelle de poids moléculaire est représentée par 'Ech 1 kb'. La flèche jaune indique la chromatine pontée et non fragmentée restée au niveau du puits du gel. Les accolades jaune et rouge indiquent l'intervalle de taille où la majorité des fragments de la chromatine sont localisés dans le cas de la chromatine soniquée pontée et dépontée, respectivement. Les valeurs à gauche du gel représentent les tailles.

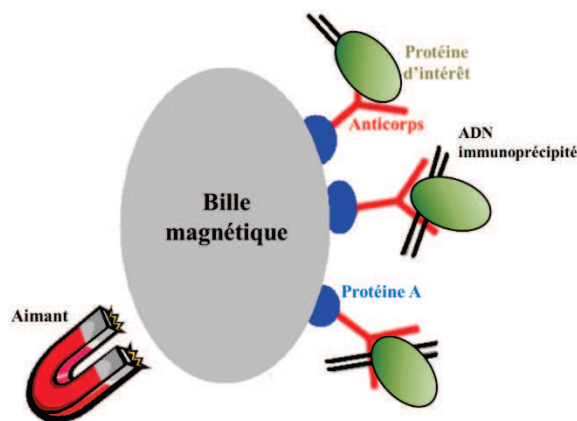


Figure II.2 : Schéma explicatif de la méthode d'immunoprécipitation avec billes magnétiques couplées à la protéine A.

L'anticorps (rouge) est lié simultanément à la protéine d'intérêt (vert foncé) en association avec l'ADN immunoprécipité (noir) et à la bille magnétique (gris) couplée à la protéine A (bleu). Le complexe est isolé à l'aide d'un aimant.

Nous avons déposé sur gel d'agarose, de la chromatine pontée non soniquée, de la chromatine pontée après sonication afin de vérifier la qualité de la sonication, notamment la taille des fragments obtenus. Nous avons également pris une petite aliquote d'échantillon sur lequel nous avons réalisé un dépontage thermique dans le but de prédire la taille des fragments ADN après la libération de l'ADN des protéines fixées dessus. Les résultats montrent que la chromatine a été effectivement pontée car on peut observer une bande intense située au niveau du puits du gel qui ne migre pratiquement pas (figII.1), bande qui correspond à la chromatine non fragmentée pontée. Le smear observé correspond à des fragments de chromatine qui appartiennent à toutes les gammes de taille, qui seraient probablement non liés à des protéines et/ou partiellement dégradés. Après sonication, on observe qu'il n'y a plus de bande au niveau du puits de gel indiquant que la chromatine a été effectivement fragmentée. On observe un smear allant de 0,2 à environ 3 kb. La partie la plus intense du smear correspond à des fragments de chromatine ayant une taille entre 0,5 et 1 kb (figII.1), ce qui correspond à une taille compatible avec la suite de l'expérimentation. Lorsqu'on réalise le dépontage, c'est-à-dire qu'on libère les liaisons ADN/ protéines de la chromatine, on observe un smear plus petit allant de 0,1 à 1,6 kb environ et dont la partie la plus intense est localisée entre 0,2 et 0,6 kb environ (figII.1). Ceci montre que l'ADN libéré a effectivement une taille inférieure aux fragments de chromatine pontée; cette taille est convenable pour la suite des analyses moléculaires.

2.1.3-Immunoprécipitation de la chromatine

L'immunoprécipitation de chromatine est réalisée selon le protocole du kit High Cell de Diagenode adapté pour une quantité de matériel de départ compatible avec ce que nous avons dans nos échantillons. Après immunoprécipitation, l'anticorps fixé sur la protéine d'intérêt (MYB221 ou MYB090) est lié à des billes magnétiques couplées à la protéine A. Le complexe billes/ protéineA / Ac / protéine / ADN (figII.2) est ensuite isolé grâce à un aimant. On libère ensuite l'ADN du complexe par chauffage. L'ADN obtenu est purifié en attendant d'être analysé par PCR ou séquencé. La quantité d'ADN immunoprécipité étant très faible, nous avons effectué un pool d'ADN provenant de plusieurs ChIP afin d'avoir une concentration d'ADN immunoprécipité suffisante pour la réalisation du séquençage.

Un échantillon d'ADN nommé Input ou "ADN contrôle" est préparé en parallèle avec chaque échantillon de ChIP. Cet échantillon correspond à de la chromatine fragmentée

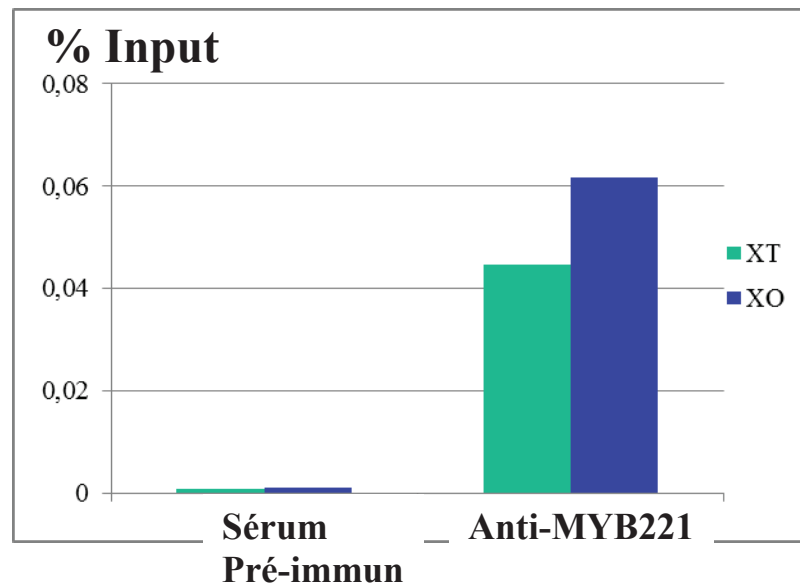


Figure II.3 : Enrichissement de l'ADN immunoprécipité par rapport à l'Input, en fragments correspondant au promoteur de la CCR.

Le niveau d'enrichissement des cibles est exprimé en % d'Input (nombre de copies IP / nombre de copies Input x100). Il est identifié par PCR quantitative en temps réel à l'aide d'amorces amplifiant une partie du promoteur de la CCR. L'immunoprécipitation a été réalisée sur des échantillons de jeune xylème de bois de tension (XT) ou de bois opposé (XO). Un contrôle négatif a été réalisé avec du sérum pré-immun.

qui subit le même traitement, sauf le passage sur les billes avec les anticorps. Il s'agit ici d'un contrôle positif représentatif de l'ADN présent avant immunoprécipitation. Il est utilisé pour déterminer les enrichissements après immunoprécipitation, que ce soit pour le séquençage haut-débit ou la PCR.

Des tests de mise au point ont été réalisés en utilisant les anticorps anti-MYB221 et des amorces correspondant au promoteur de la CCR ou de la CAD. Nous faisons l'hypothèse que comme EgMYB1, MYB221 se fixe sur ces 2 promoteurs. Un contrôle négatif a été réalisé avec du sérum pré-immun; l'immunoprécipitation a été réalisée à la fois sur du jeune xylème de bois de tension et de bois opposé (figII.3). On mesure un enrichissement quel que soit l'échantillon pour le promoteur CCR, de 50 à 60 par rapport à l'échantillon contrôle (sérum pré-immun), mais cet enrichissement reste faible en % d'input. Pour le promoteur CAD, nous n'avons pas mesuré d'enrichissement. Les résultats ont été jugés satisfaisants pour passer à l'étape ChIP-SEQ.

2.2- Résultats obtenus pour MYB221

2.2.1- Résultats de séquençage et d'analyses primaires pour MYB221

L'immunoprécipitation de la chromatine a été réalisée sur du jeune xylème tendu à l'aide d'anticorps anti-MYB221. Un premier essai a été conduit pendant l'année 2012. L'ADN immunoprécipité provenant de 6 expériences ChIP (IP-MYB221-A) a été poolé afin d'obtenir une quantité d'ADN suffisante (5 ng) permettant la préparation des banques pour le séquençage haut débit. Un ADN contrôle nommé Input-A a été également préparé en même temps. Une répétition biologique a été réalisée en 2013 sur le jeune xylème tendu d'une autre série de plantes : on a poolé l'ADN immunoprécipité de 12 expériences ChIP (IP-MYB221-B) avant le séquençage. Un échantillon d'ADN contrôle nommé Input-B a été également préparé. Les banques ont été préparées à partir de 3 à 10 ng d'ADN immunoprécipité ou contrôle.

Après séquençage, les échantillons ont passé avec succès l'ensemble des tests de qualité requis par la plateforme de séquençage (Annexe 6). Les séquences des tags de 36 pb ont été alignées avec la troisième version du génome de peuplier (v3.0) qui venait d'être disponible (*Populus trichocarpa* v3.0, DOE-JGI, <http://www.phytozome.net/poplar>). C'est

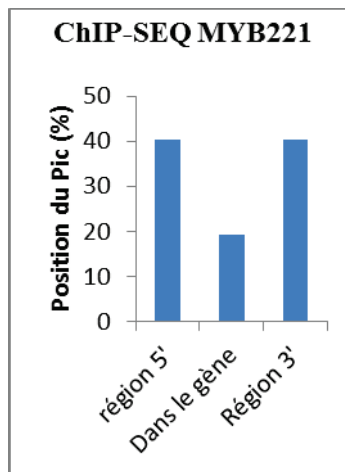


Figure II.4 : Position des pics par rapport aux modèles de gène à proximité.

La position des pics par rapport aux gènes à proximité est exprimée en % par rapport au nombre total de gènes cibles communs.

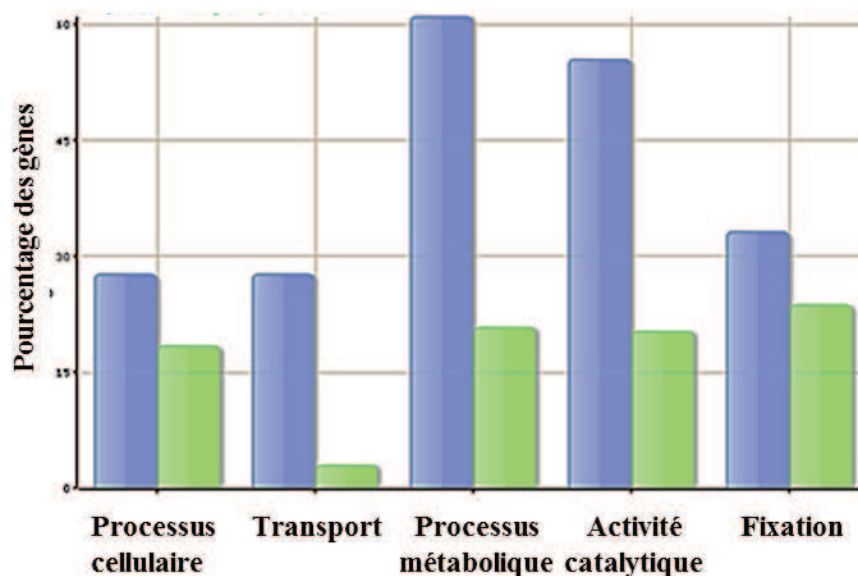


Figure II.5 : Classification des cibles de MYB221 selon leur fonction moléculaire ou leur implication dans un mécanisme cellulaire.

Le classement est fondé sur l'annotation Gene Ontology (GO). Les barres bleues correspondent aux cibles potentielles de MYB221 exprimées en % par rapport au nombre total de gènes testés. Pour chaque catégorie, ce pourcentage est comparé à une référence (pourcentage de gènes testés au sein d'un génome de référence). Schéma réalisé à l'aide de l'outil agriGO.

pourquoi les séquences du premier essai de 2012 ont été réalignées sur cette version pour avoir des résultats comparables pour les deux essais.

La détection des pics d'enrichissement se fait par utilisation du logiciel CisGenome (Ji *et al.* 2008): on obtient un set de pics par couple d'échantillons (IP-MYB221-A/Input-A et IP-MYB221-B/Input-B). N'ont été retenus que les pics communs aux 2 répétitions biologiques. 66 pics communs, de 200 pb, ont été identifiés pour lesquels 56 gènes ont été identifiés à 3 kb de part et d'autre des pics (Annexe 8).

Une première classification générale a été faite selon la position de la médiane de chaque pic par rapport aux gènes à proximité des pics. Pour MYB221, les résultats montrent une distribution égale des pics entre la région 5' et la région 3' des gènes cibles d'environ 40 %. Les pics "dans le gène" situés en partie dans la séquence génomique des modèles de gènes représentent 20 % environ du total des gènes cibles (figII.4). Pour une partie d'entre eux, la médiane du pic est située en partie en 5'UTR ; le pic est comptabilisé comme étant "dans le gène".

2.2.2- Classification fonctionnelle des cibles potentielles de MYB221

Nous avons procédé à une classification fonctionnelle des gènes cibles potentiels de MYB221 selon leurs identifiants GO (Gene Ontology). Les cibles dont le pic est situé du côté 3' des gènes cibles n'ont pas été prises en compte dans un premier temps, car on suppose que ce sont de faux positifs puisque les zones promotrices se situent généralement du côté 5'. Les résultats montrent que plus de 60 % des cibles sont des gènes impliqués dans le processus métabolique et/ou l'activité catalytique (figII.5). Les cibles ayant des fonctions de transport représentent plus d'un quart du total des cibles. Cependant, il est à noter que pour MYB221, nous n'avons obtenu qu'un très faible nombre de pics communs (66), ce qui rend la classification fonctionnelle relativement peu informative.

2.2.3- Recherche de motifs consensus dans les séquences des pics pour MYB221

De même que pour la classification fonctionnelle des gènes cibles, seuls les pics en position 5' et "dans le gène" sont pris en compte pour la recherche des motifs consensus. Les résultats montrent qu'un motif de 10 nucléotides (CCTACCTACC) est bien conservé entre les pics des deux répétitions biologiques (figII.6-A, C) et qui est présent au moins sur la moitié

des séquences de pics testées (figII.6-B, D). Au sein de ce motif on peut identifier le motif SMRE 8 (Zhong et Ye 2012). Deux autres variantes du motif SMRE (6 et 2) ont été trouvées dans les autres motifs identifiés (figII.6-A, C).

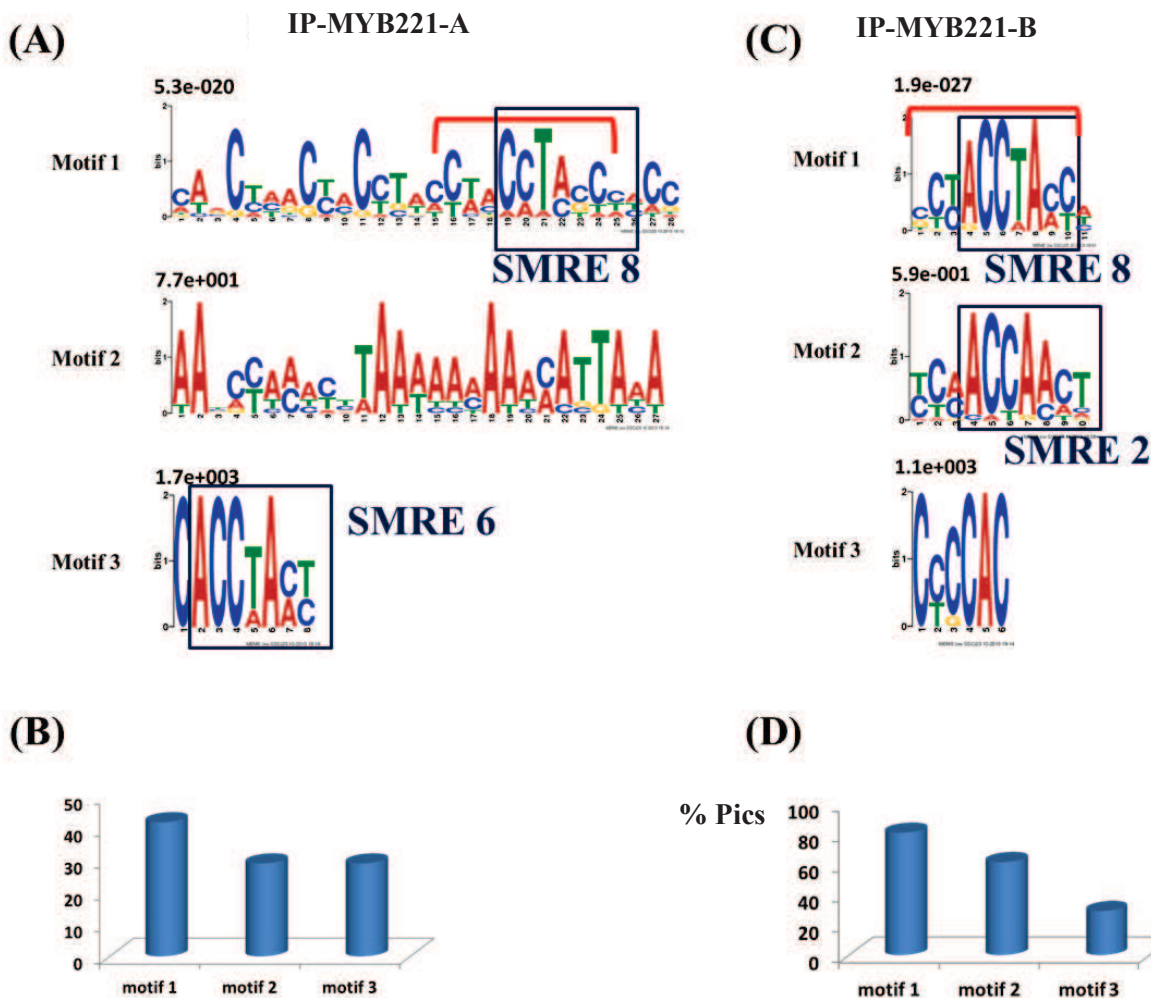


Figure II.6 : Recherche des motifs consensus dans les séquences ADN des pics correspondant aux zones promotrices potentielles de gènes cibles de MYB221. Les trois meilleurs motifs identifiés sont présentés pour chaque échantillon d'ADN immunoprécipité : IP-MYB221-A (A) et IP-MYB221-B (C). La E value est notée au-dessus de chaque motif. B et D : pourcentage de répartition de chaque motif par rapport au nombre total des séquences testées (ici séquences de 31 pics testées correspondant aux pics en position 5' ou ''dans le gène''). Les motifs sont analysés par l'outil MEME-ChIP. Le motif encadré par un rectangle correspond au motif SMRE : ACC[A/T]A[A/C][T/C] (Zhong et Ye 2012) avec ses variantes 2, 6 et 8. La partie du motif 1 indiquée par une accolade rouge correspond à une partie identique entre IP-MYB221-A et IP-MYB221-B. Les motifs MEME sont représentés par des matrices de probabilité de position spécifiques qui indiquent la probabilité de chaque lettre apparaissant à chaque position possible dans une occurrence du motif. Ceux-ci sont affichés sous forme de "logos de séquences", contenant des piles de lettres à chaque position dans le motif. La hauteur totale de la pile est le "contenu de l'information" de cette position dans le motif en bits. La hauteur des lettres individuelles d'une pile est la probabilité de la lettre à cette position multipliée par le contenu de l'information totale de l'empilement.

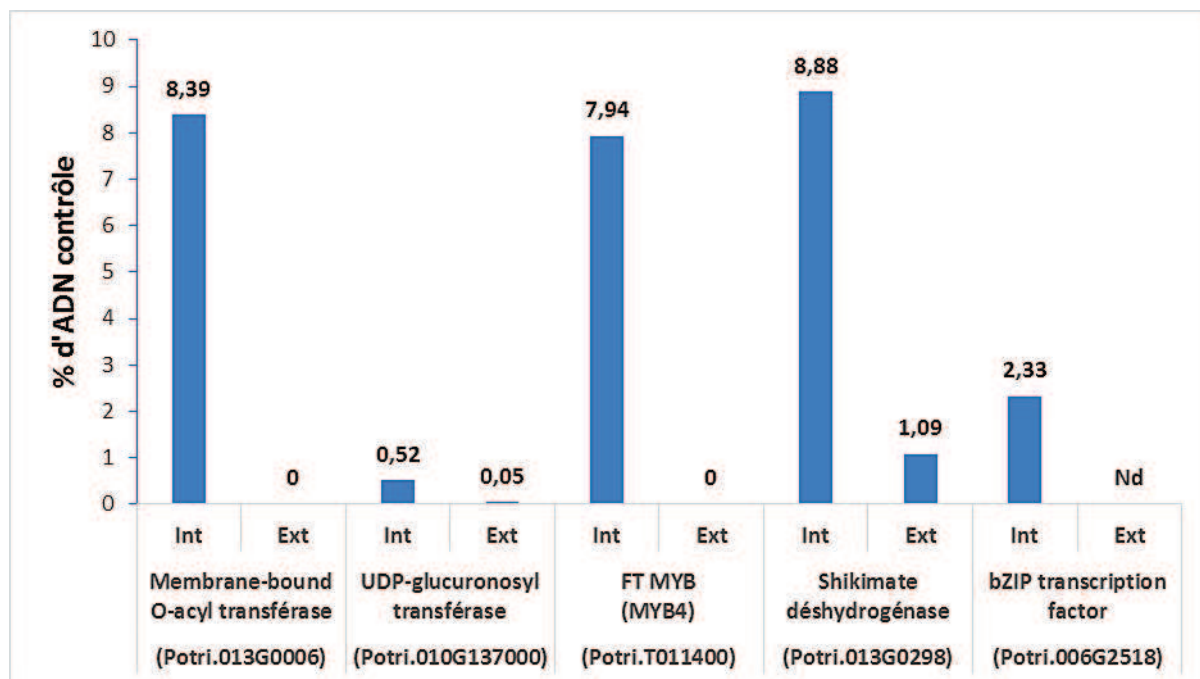


Figure II.7 : Enrichissement de l'ADN immunoprécipité pour 5 zones de fixation potentielles de MYB221, identifiées par ChIP-SEQ.

Le niveau d'enrichissement des cibles est exprimé en % d'ADN contrôle (nombre de copies ADN immunoprécipité / nombre de copies ADN contrôle (Input) x100). Il est mesuré par PCR quantitative en temps réel à l'aide d'amorces amplifiant dans la séquence du pic considéré (Int) et à 1kb du pic (Ext), et à partir de courbes de calibration réalisées avec de l'ADN génomique de 101-74. Nd : non déterminé, car pas d'amplification mesurée (amorces non fonctionnelles).

2.2.4- Validation de 5 cibles de MYB221 par ChIP-qPCR

La technique ChIP-SEQ donne comme résultat une liste de cibles directes potentielles au sein du génome entier, la zone de fixation du FT étant incluse dans le pic correspondant. Nous avons cherché à vérifier qu'on retrouve bien les mêmes pics enrichis après ChIP-qPCR en utilisant des amorces pouvant amplifier la séquence ADN du pic, ainsi qu'un couple d'amorces amplifiant une séquence à 1kb en amont du pic (contrôle négatif). Nous avons utilisé des échantillons provenant d'une troisième répétition biologique indépendante : des échantillons immunoprécipités avec l'anticorps anti-MYB221, et l'ADN contrôle. Les 5 gènes cibles ont été choisis en prenant comme critères leur annotation, leur position par rapport au pic issu de Chip-SEQ (< 1,5 kb), le degré de significativité du pic associé (déterminé par CisGenome) et la présence *in silico* des sites SMRE. Ces gènes sont : (1) un gène codant pour un FT MYB similaire à AtMYB4 et qui contient un motif de répression de type LxLxLx, comme MYB221 et MYB156, (2) un gène codant pour une O-acyl transférase associée à la membrane (MBOAT), une enzyme qui joue un rôle important dans le métabolisme des lipides chez les plantes, (3) un gène codant pour un FT de type bZIP, (4) un gène codant pour une shikimate déshydrogénase qui catalyse une étape de la voie du shikimate qui peut donner lieu à des composants chimiques impliqués par la suite dans la voie de biosynthèse de lignines, (5) un gène codant pour une UDP-glucuronosyl transférase responsable du processus de glucuronidation. Les résultats sont exprimés en % de cibles détectées par rapport aux cibles présentes dans l'ADN contrôle, avant immunoprécipitation (figII.7). A partir de courbes de calibration réalisées à partir d'ADN génomique de 101-74, on a une mesure quantitative du nombre de copies de l'amplicon pour chaque couple d'amorces. Les résultats montrent que l'ADN immunoprécipité est fortement enrichi pour 3 pics (autour de 8%) (FT MYB, MBOAT et Shikimate déshydrogénase) et plus faiblement pour le pic correspondant au FT bZIP (2%). Le pic correspondant à l'UDP-glucuronosyl transférase ne semble pas enrichi (figII.7). Cet enrichissement est spécifique de la zone de fixation de MYB221 sur les ADN cibles car il n'est observé qu'au niveau des pics, et non pas en dehors des pics : quand on s'éloigne de ces derniers de 1 kb, on n'observe plus d'enrichissement (figII.7). Une 6^{ème} cible avait été retenue, une alcool déshydrogénase, mais les deux couples d'amorces choisies, que ce soit dans le pic ou en dehors du pic, n'ont pas permis d'amplifier de séquence. Donc, on ne peut pas conclure pour ce gène cible.

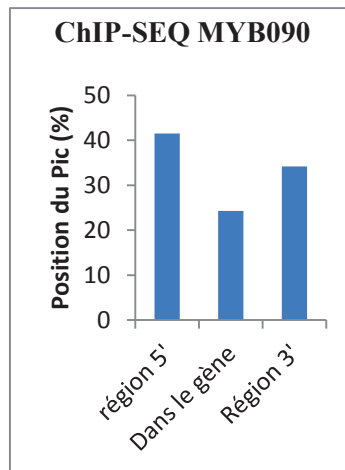


Figure II.8 : Position des pics par rapport aux modèles de gène à proximité.

La position des pics par rapport aux gènes cibles de MYB090 est exprimée en % par rapport au nombre total de gènes cibles communs.

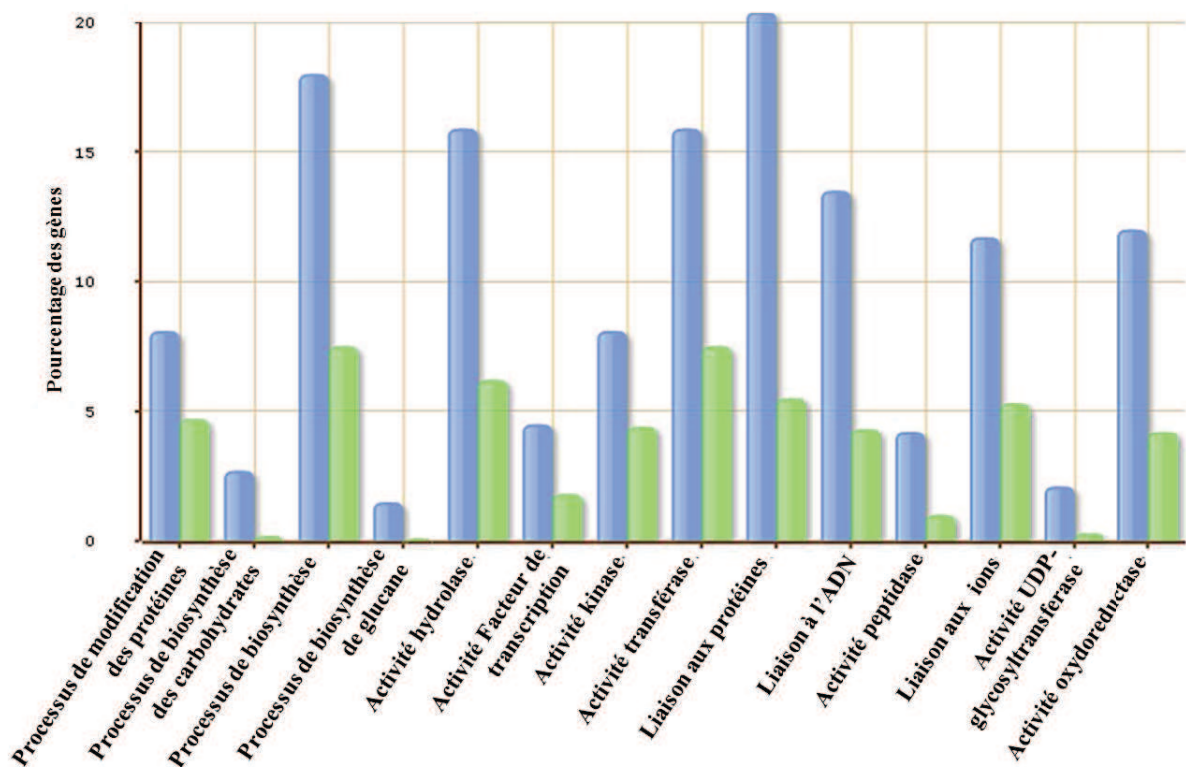


Figure II.9 : Classification des cibles de MYB090 selon leur fonction moléculaire ou leur implication dans un mécanisme cellulaire.

Le classement est fondé sur l'annotation Gene Ontology (GO). Les barres bleues correspondent aux cibles potentielles de MYB090 exprimés en % par rapport au nombre total de gènes testés. Pour chaque catégorie, ce pourcentage a été comparé à une référence (pourcentage de gènes testés au sein d'un génome de référence, celui de *Populus trichocarpa* v2.0). Schéma réalisé à l'aide de l'outil agriGO.

2.3- Résultats obtenus pour MYB090

2.3.1- Résultats de séquençage et d'analyses primaires pour MYB090

L'immunoprécipitation de la chromatine a été réalisée sur du jeune xylème tendu à l'aide d'anticorps anti MYB090. L'ADN immunoprécipité provenant de 18 expériences ChIP a été poolé. Deux essais ont été réalisés en 2013 (IP-MYB090-A et IP-MYB090-B) avec un ADN contrôle nommé Input-C et le séquençage a été fait par multiplexage des trois échantillons d'ADN sur la même "flow cell". Les banques sont préparées à partir de 10 ng d'ADN. Après séquençage, l'ensemble des échantillons a passé avec succès l'ensemble des tests de qualité requis par la plateforme de séquençage (Annexe 7). Les séquences des tags ont été alignées avec la troisième version du génome de peuplier (v3.0). Nous avons obtenu 901 pics communs entre les deux échantillons d'ADN immunoprécipité qui correspondent à 874 gènes cibles, à 3 kb de part et d'autre des pics (Annexe 9).

La classification générale faite selon la position des pics par rapport aux gènes cibles montre une distribution similaire des pics entre la région 5' et la région 3' des gènes cibles. Les pics situés dans les gènes représentent environ ¼ du total des gènes cibles (figII.8).

2.3.2- Classification des cibles de MYB090

Nous avons procédé à une classification des gènes cibles de MYB090 selon leurs identifiants GO (Gene Ontology) en ne prenant en compte que les gènes dont le pic est détecté du côté 5' ou bien "dans le gène". Nous avons ici 334 cibles annotées qui sont analysées alors que 259 ne le sont pas et donc non sujettes à cette analyse. Chaque catégorie est plus enrichie en gènes cibles par rapport à la référence (figII.9). Les résultats montrent aussi une diversité importante des cibles testées (figII.9). Une grande partie de ces cibles correspond à des gènes ayant une fonction de liaison (plus de 20 %), de modification (7 à 8 %) et de dégradation (5 %) des protéines. Aussi, près de 16 % des cibles sont des transférases parmi lesquelles on trouve des glycosyltransférases (GT). Un nombre faible de gènes (environ 2 %) codant pour des UDP-glycosyltransférases, notamment des gènes de cellulose synthase, a été identifié. De plus, environ 2 à 3 % des cibles sont impliquées dans la biosynthèse des carbohydrates (comprenant la cellulose et d'autres glucanes). De même, bien qu'environ 14 % des gènes cibles ont des fonctions de liaison à l'ADN, seulement 5 % ont une activité de FT.

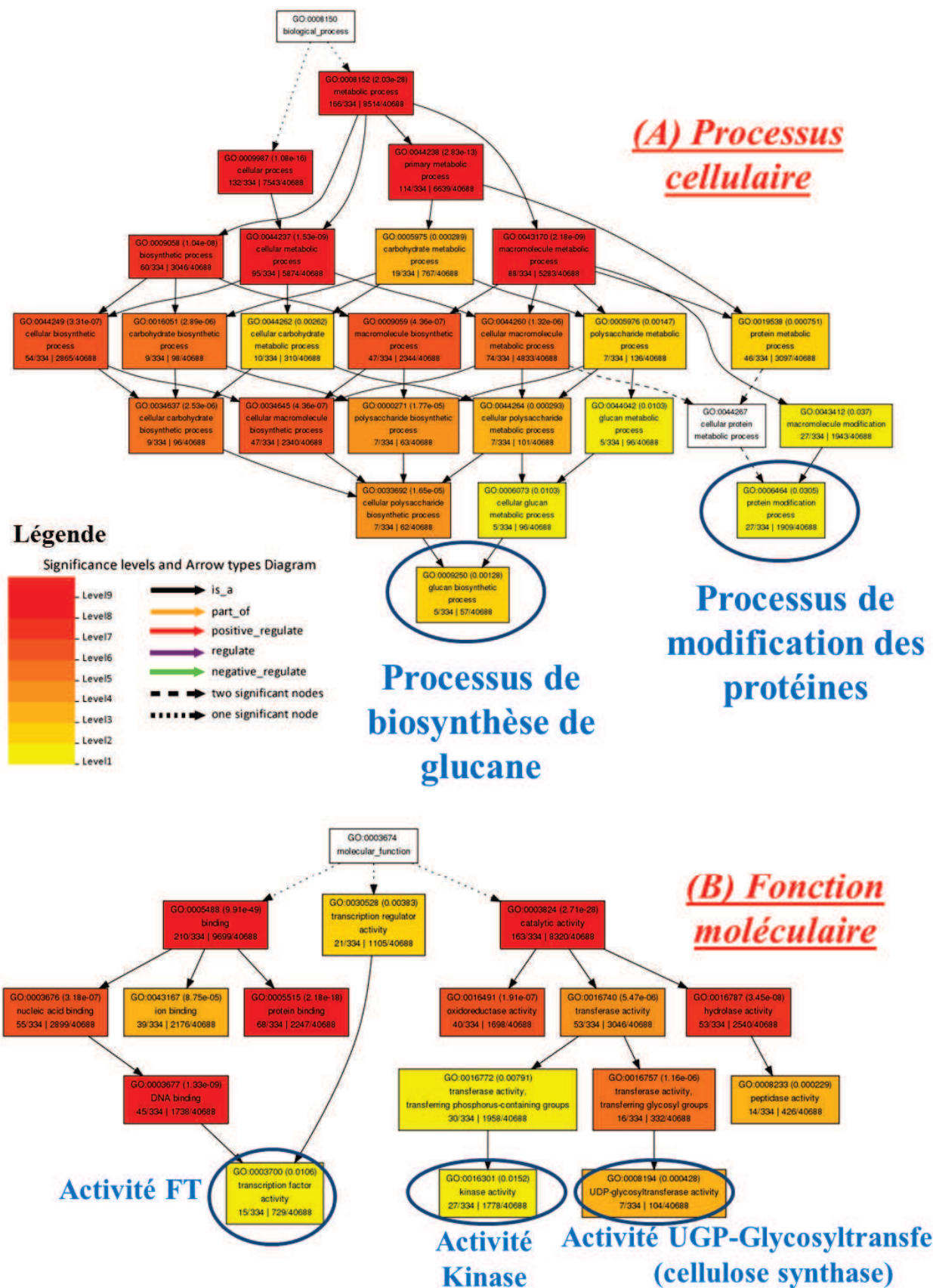
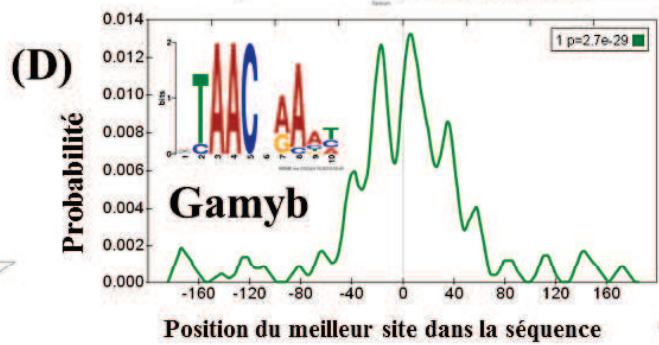
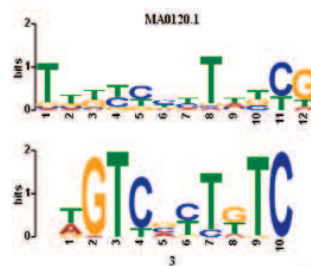
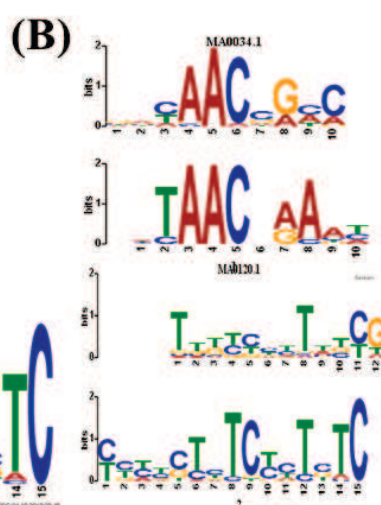
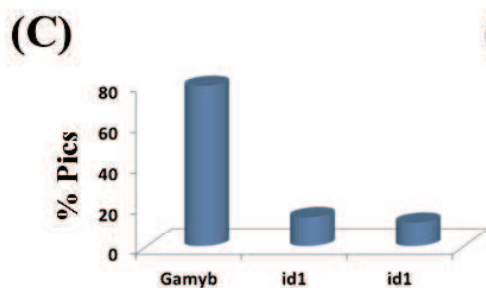
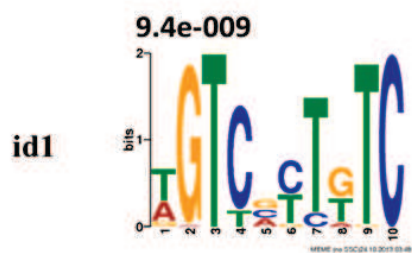
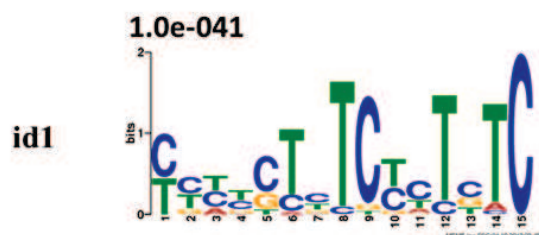
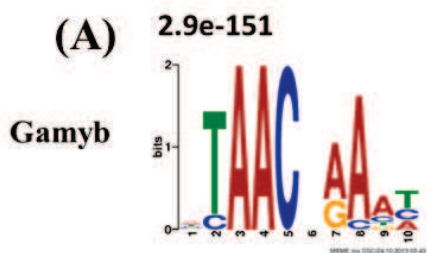


Figure II.10 : Présentation graphique hiérarchique obtenue après classification des cibles de MYB090 selon (A) : le processus ou mécanisme cellulaire et (B) : la fonction moléculaire.

Graphique dessiné à l'aide de l'outil agriGO.

IP-MYB090-A



Name	MA0034.1
Alt. Name	Gamyb
Database	JASPAR_CORE_2009 _plants.meme
p-value	0.0094295
E-value	0.19802
q-value	0.344142
Overlap	9
Offset	1
Orientation	Normal

Name	MA0120.1
Alt. Name	id1
Database	JASPAR_CORE_2009 _plants.meme
p-value	0.00126826
E-value	0.0266334
q-value	0.0532668
Overlap	11
Offset	-4
Orientation	Normal

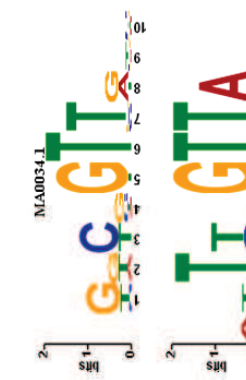
Name	MA0120.1
Alt. Name	id1
Database	JASPAR_CORE_2009 _plants.meme
<i>p</i>-value	0.000841148
<i>E</i>-value	0.0176641
<i>q</i>-value	0.0353282
Overlap	10
Offset	1
Orientation	Normal

IP-MYB090-B

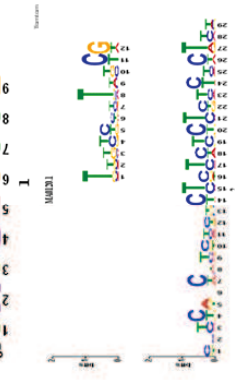
(E)

(F)

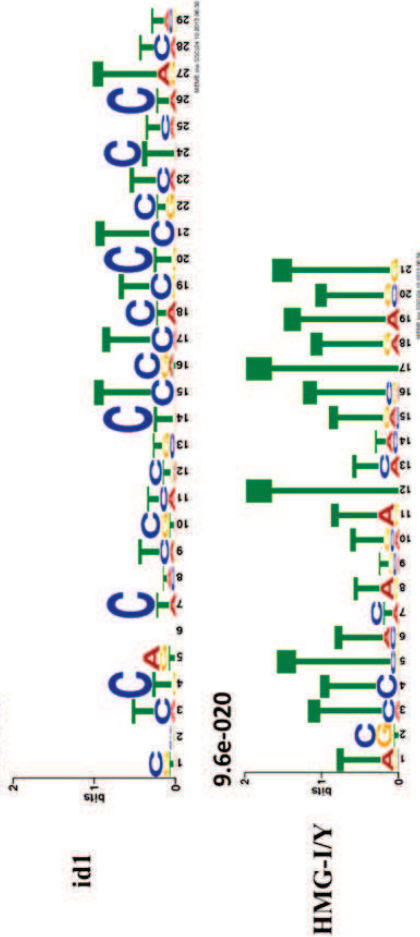
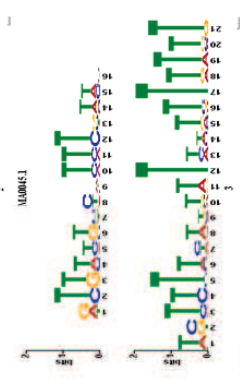
Name	MA0034.1
Alt. Name	Gamyb
Database	JASPAR_CORE_2009_plants.meme
p-value	0.0301413
E-value	0.632968
q-value	0.632968
Overlap	8
Offset	-1
Orientation	Reverse



Name	MA0120.1
Alt. Name	id1
Database	JASPAR_CORE_2009_plants.meme
p-value	0.0132623
E-value	0.278507
q-value	0.278507
Overlap	12
Offset	-15
Orientation	Normal

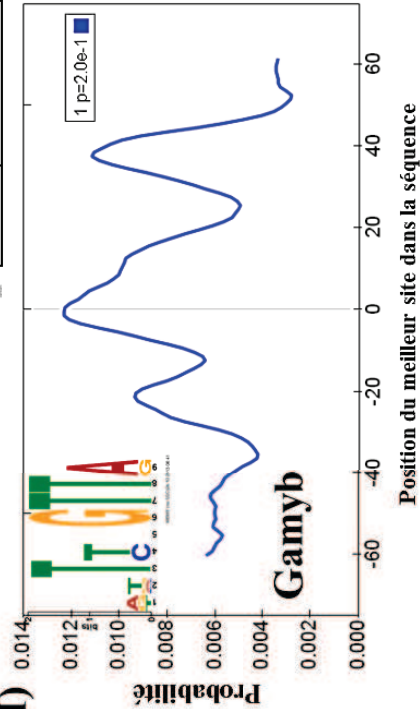
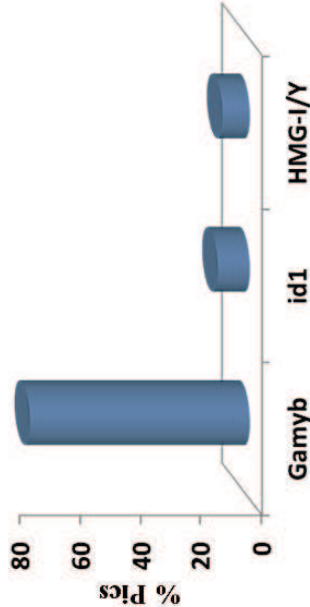


Name	MA0045.1
Alt. Name	HMG-I/Y
Database	JASPAR_CORE_2009_plants.meme
p-value	0.00246435
E-value	0.0517514
q-value	0.103503
Overlap	16
Offset	-2
Orientation	Reverse



(G)

(H)



Des enzymes diverses comme les hydrolases, les oxydoréductases et les kinases représentent environ 35 % des cibles (figII.9).

Une deuxième présentation graphique réalisée en se basant sur les termes GO des gènes cibles de MYB090 en les classant de façon hiérarchique suggère qu'au niveau moléculaire, MYB090 aurait un rôle important dans la régulation de l'activité UDP-glycosyltransférase notamment via des cellulose synthases. Il jouerait aussi un rôle dans la régulation des kinases et de l'activité de certains facteurs de transcription (figII.10-B). Au niveau cellulaire, le graphique propose que les cibles de MYB090 s'organisent pour aboutir à deux fonctions principales : la biosynthèse des glucanes et la modification des protéines (figII.10-A).

2.3.3- Recherche de motifs consensus dans les séquences des pics pour MYB090

De même que pour la classification fonctionnelle des gènes cibles, seuls les pics en position 5' et ''dans le gène'' sont pris en compte pour la recherche des motifs consensus. Pour chaque échantillon d'ADN immunoprécipité, 3 motifs potentiels conservés ont été identifiés (figII.11-A, E) dont un motif semblable à Gamyb (connu pour être impliqué dans la réponse des plantes aux gibbérellines) et dont la séquence est [ATC]TAAC[ATG][AG]-AA[TCA]. Ce motif est conservé dans la majorité des séquences promotrices cibles pour les deux échantillons d'ADN (figII.11-C, G) et semble être enrichi au centre de celles-ci (figII.11-D, H). Ceci indique que ce motif pourrait avoir un rôle déterminant dans la régulation des gènes cibles par MYB090. Pour une petite partie de séquences cibles (moins de 20 %) (figII.11-C, G), un autre motif conservé a été identifié comme similaire au motif id1 (inhibitor of DNA binding 1).

Figure II.11 : Recherche des motifs consensus dans les séquences ADN des pics correspondant aux zones promotrices potentielles des gènes cibles de MYB090. Les trois meilleurs motifs identifiés sont présentés pour chaque échantillon d'ADN immunoprécipité : IP-MYB090-A et IP-MYB090-B. La E value est notée au-dessus de chaque motif. C et G : pourcentage de répartition de chaque motif par rapport au nombre total des séquences testées (ici séquences de 521 pics testées correspondant aux pics en position 5' ou ''dans le gène''). D et H : enrichissement des séquences des pics en meilleur motif trouvé pour chaque échantillon. B et F : alignement des motifs trouvés avec le motif le plus proche chez les plantes (JASPAR-CORE-2009-PLANTS). Les données relatives à chaque motif sont présentées dans les tableaux en B et F. Les motifs sont analysés par l'outil MEME-ChIP. Les motifs MEME sont représentés par des matrices de probabilité de position spécifiques qui indiquent la probabilité de chaque lettre apparaissant à chaque position possible dans une occurrence du motif. Ceux-ci sont affichés sous forme de "logos de séquences", contenant des piles de lettres à chaque position dans le motif. La hauteur totale de la pile est le "contenu de l'information" de cette position dans le motif en bits. La hauteur des lettres individuelles d'une pile est la probabilité de la lettre à cette position multipliée par le contenu de l'information totale de l'empilement.

3- Discussion

L'étude de la fonction des FT chez les plantes peut être basée sur des approches de génétique inverse (transgénèse et mutants) qui permettent d'identifier des gènes différentiellement régulés dans des plantes transgéniques ou mutants : ces approches ne permettent pas de savoir si cette régulation *via* le FT est directe ou indirecte. La technique ChIP-SEQ permet l'identification des cibles directes des FT, sans *a priori* sur l'identité de ces cibles, ce qui permet la découverte de nouvelles cibles. C'est une technique *in vivo*, le pontage se faisant très rapidement après le prélèvement pour fixer les complexes de régulation en place sur la chromatine, elle est donc plus informative que certaines techniques *in vitro* comme le simple hybride ou le gel retard. Par contre, elle est fondée sur l'utilisation d'anticorps, qui doivent être spécifiques et à fort pouvoir immunoprécipitant. Dans ce travail de thèse, nous avons utilisé cette technique pour identifier les cibles ADN de deux FT, MYB221 et MYB090 à l'échelle du génome du peuplier. Cette technique est relativement nouvelle chez les plantes et n'a jamais été, jusque-là, publiée sur les arbres.

Dans notre étude, nous avons réussi à mettre au point la technique ChIP-SEQ sur du jeune xylème de peuplier, qui possède des parois cellulaires épaisses. Le pontage avec infiltration sous-vide permet au formaldéhyde de pénétrer ces parois jusqu'au noyau. Notre protocole expérimental permet de ponter, sans avoir de problèmes de sur-fixation de la chromatine, car nous pouvons soniquer la chromatine après pontage pour obtenir des fragments de taille désirée pour la suite de l'expérience (figII.1, entre 0,2 et plus de 3 kb avec une majorité de fragments entre 0,5 et 1 kb). Ceci est similaire aux données de la littérature (Bowler *et al.* 2004). Il faut noter cependant que toutes les cellules ne sont vraisemblablement pas pontées, ce qui peut expliquer le smear observé dans le puits « Chromatine pontée, non soniquée » (figII.1) : nous travaillons avec une population de cellules, dont certaines sont pontées et d'autres non. Nous avons aussi vérifié par ChIP-qPCR que l'immunoprécipitation fonctionne et donne lieu à un enrichissement effectif en cibles (figII.3, figII.7). Dans le chapitre 1, nous avons vu que l'anticorps MYB221 pouvait reconnaître la protéine recombinante MYB090, alors que l'anticorps MYB090 ne reconnaît pas la protéine recombinante MYB221. Parmi les cibles de MYB221, nous ne retrouvons que 11 cibles potentielles de MYB090 (soit 1,3% du total de cibles de MYB090) dont 6 cibles non annotées, une hydrolase, une aquaporine PIP, un FT bZIP et deux autres en liaison avec le métabolisme des lipides et l'import de protéines. Ceci montre que l'anticorps n'a pratiquement pas immunoprécipité des cibles de MYB090, dans des conditions *in vivo* de

protéines natives. Cependant, on ne peut pas exclure que l'un ou l'autre ait précipité d'autres FT MYB, proches au niveau séquence et structure protéique.

Quel que soit l'anticorps considéré, environ 40 % des pics détectés sont situés du côté 3' des cibles potentielles identifiées. Cette valeur est relativement similaire à ce qui a été rapporté dans Brandt *et al.* (2012) qui ont trouvé environ 30 % de pic en 3' de gènes dans une étude ChIP-SEQ sur *Arabidopsis*. Ce pourcentage semble important car on s'attend plutôt à détecter les pics dans la zone promotrice des cibles potentielles situées en région 5' ou sur des introns dans la région "dans le gène". Cela peut s'expliquer de différentes façons : (1) ces pics sont de faux positifs. Il y a effectivement un risque que le formaldéhyde puisse fixer des interactions ADN-protéines transitoires interférant ainsi dans la détection des interactions stables, comme il peut causer aussi l'association indirecte des protéines avec l'ADN (Das *et al.* 2004). Comme on peut supposer que les faux-positifs sont répartis de façon homogène le long du génome, et donc qu'on a également des faux-positifs en 5' ou "dans le gène", cela implique que les cibles soient validées par des techniques complémentaires, comme des tests d'expression transitoire permettant de suivre la répression ou l'activation d'un gène rapporteur en présence du FT étudié ; (2) certains, au moins, de ces pics ne sont pas de faux-positifs. Chez les mammifères notamment, il a été montré que les séquences enhancer pouvaient se trouver en région 3' de gènes, voire à très grande distance du locus considéré (plusieurs dizaines de kb, Bulger et Groudine 2010). La structure tridimensionnelle de la chromatine permet de mettre en relation des protéines qui se fixent sur des zones du génome relativement éloignées. Dans un premier temps, nous n'avons pas pris en compte ces pics détectés en 3' des gènes et nous avons affecté le gène le plus proche du pic (moins de 3 kb) comme cible potentielle du FT.

Le nombre de pics obtenus pour chaque expérience ChIP-SEQ pour MYB221 et MYB090 (IP-MYB221-A = 919 pics ; IP-MYB221-B = 1738 pics ; IP-MYB090-A : 6555 pics, IP-MYB090-B : 1000 pics) est élevé surtout pour IP-MYB090-A. Ce nombre peut paraître élevé et souligne la complexité des réseaux de régulation transcriptionnelle. Le nombre de pics communs (66) pour MYB221 est largement plus faible que celui des pics communs pour MYB090 (901). Ceci pourrait être expliqué par le fait que (1) les deux expériences ChIP réalisées pour MYB221 sont moins homogènes que celles réalisées pour MYB090, car elles ont portées sur des expérimentations décalées dans le temps (2012 et 2013), pour lesquelles il a été plus difficile de maintenir des conditions de croissance identiques ; (2) MYB090 a physiologiquement plus de cibles que MYB221.

La recherche de motifs consensus dans les séquences des pics pour chaque FT a montré que les deux expériences se distinguent par des motifs différents. Les cibles potentielles de MYB221 identifiées par ChIP-SEQ ont un motif SMRE (Zhong et Ye 2012) conservé dans la moitié de séquences de pics correspondant aux zones de fixation de MYB221 dans les régions promotrices des gènes cibles. De même, un motif similaire au domaine Gamyb a été conservé chez la plupart des séquences des pics communs de deux expériences ChIP-SEQ sur MYB090. Ce motif a présenté une distribution plutôt centrale par rapport à la séquence des pics. Ceci est en accord avec le principe de la technique : le FT, au sein d'un complexe multiprotéique, va protéger l'ADN cible de la sonication, et donc on génère une population de fragments d'ADN à séquencer, qui vont donner le pic, dont la médiane correspond au site de fixation du FT sur l'ADN. Le domaine Gamyb est un domaine de fixation à l'ADN connu pour le FT GAMYB, un FT du groupe R2R3MYB qui a un rôle dans la signalisation par les gibbérellines et la régulation de la plupart des gènes inductibles par les gibbérellines (Kaneco *et al.* 2004, Aya *et al.* 2009).

Les cibles identifiées de MYB221 ont principalement des fonctions liées au métabolisme, au transport, à la fixation à l'ADN et à d'autres processus cellulaires. On trouve par exemple un FT similaire à MYB4 (Jin *et al.* 2000) qui pourrait être un répresseur de la lignification, un gène codant pour une shikimate déshydrogénase qui catalyse une étape de la voie du shikimate, qui peut donner lieu à des intermédiaires de la voie de biosynthèse de lignines (Bonawitz et Chapple 2010), un gène codant pour une UDP-glucuronosyl transférase responsable du processus de glucuronidation. Le gène CAD a été identifié comme cible de MYB221 dans les deux échantillons de ChIP-SEQ mais n'est pas apparu dans la liste des cibles communes car les pics correspondants ne sont pas superposés. De même, le gène CCR apparaît dans la liste des cibles de MYB221 mais uniquement au niveau du deuxième échantillon (IP-MYB221-B) à côté de 3 pics différents. Ceci, d'un côté, laisse à revoir la méthode de sélection des cibles pour MYB221. Pourquoi ne pas regarder en plus des pics communs, les gènes communs ? Pour MYB090, les fonctions des cibles potentielles sont assez diversifiées. Certains sont des FT de plusieurs types (MYB, ZFP, bHLH ...), d'autres sont des enzymes dont certaines seraient impliquées dans la biosynthèse de cellulose et de lignines. D'ailleurs, 3 gènes codant pour une CCoAOMT ont été identifiés parmi les cibles potentielles de MYB090. Ceci est en concordance avec l'hypothèse de Karpinska *et al.* (2004) suggérant qu'un gène de CCoAOMT est régulé par PttMYB21a, l'orthologue de MYB090 chez *Populus tremula x tremuloides*. D'autres sont des glycosyltransférases (GTs) qui ont un

rôle dans la formation de la paroi secondaire au niveau de la synthèse des polysaccharides pendant la formation du bois (Brown et al. 2005). Ici, nous avons identifié comme cible potentielle GT8 impliquée dans la biosynthèse des xylanes (pour revue voir Mellerowicz et Sundberg 2008). De plus, plusieurs gènes potentiellement impliqués dans la signalisation avec les phytohormones notamment les auxines, les cytokinines et les gibbérellines ont été identifiés. De même, nous avons trouvé plusieurs cibles potentielles codant pour des protéines à arabinogalactanes à domaine fascicline (FLAs).

Chapitre 3
Fonction *in planta*
des facteurs de transcription

Pour mieux comprendre la/les fonction(s) de MYB221, MYB156 et MYB090 dans la plante, nous avons utilisé la transgénèse, outil permettant de relier la modification de l'expression d'un gène avec la modification du phénotype de la plante. Nous avons transformé des plants de peuplier selon deux approches différentes. La première approche a consisté en la sur-expression de la séquence codante de chaque FT, sous le contrôle d'un promoteur constitutif fort dans des plants de peuplier afin d'en déterminer les éventuelles modifications résultantes. La deuxième approche a consisté en la surexpression de la séquence du FT fusionnée, au niveau de son extrémité 3', à une séquence de 42 nucléotides correspondant, au niveau protéique, à un motif de répression dominante (SRDX) (Hiratsu *et al.* 2003). Après la fixation du FT sur l'ADN de ses cibles à l'aide de son domaine de fixation, ce motif entraînera la répression de la transcription des séquences cibles. Grâce au promoteur fort, la quantité produite de FT fusionné au domaine SRDX (FT-SRDX) devrait être largement supérieure à celle du FT endogène. Ainsi, il y aura une concurrence 'inégal' entre le FT-SRDX et ce dernier pour se fixer sur les zones d'ADN cibles. L'expression des gènes cibles du FT sera alors réprimée et on pourra avoir des éléments sur la/les fonction(s) du FT *in planta* si ce dernier est un activateur de transcription des gènes cibles.

1- Production des lignées transgéniques

Les vecteurs de transformation ont été préparés à partir du vecteur pMDC32 (Curtis et Grossniklaus, 2003) permettant la sur-expression de gène sous le contrôle du promoteur constitutif fort CaMV 35S doublé. Les séquences correspondantes à la séquence codante des 3 FT étudiés, fusionnées ou non au motif SRDX, ont été clonées par recombinaison selon la technologie Gateway (voir Matériels et Méthodes). Les vecteurs binaires ont été mobilisés dans des agrobactéries qui, ont été utilisées pour transformer des explants de peuplier. Subséquemment, avec 3 FTs et 2 approches, nous avons réalisé 6 transformations qui ont abouti à des cals transgéniques sur-exprimant le FT (MYB221 ; MYB156 ; MYB090) ou le FT-SRDX (MYB221-SRDX ; MYB156-SRDX ; MYB090-SRDX).

Pour chaque transformation, un nombre maximal de 60 cals transgéniques différents a été isolé à partir des explants transformés, chaque cal étant considéré comme un événement indépendant de transformation. Ensuite, ces cals ont été mis en culture sur un milieu favorisant la production de micro-tiges. Ils ont été divisés à chaque fois que leur taille devenait trop importante. Les micro-tiges provenant d'un même cal ont été considérées comme identiques,

Tableau III.1 : Nombre de lignées transgéniques isolées par transformation

<i>Nom de la transformation</i>	<i>Nombre de lignées</i>
MYB221 OE	24
MYB221-SRDX OE	24
MYB156 OE	22
MYB156-SRDX OE	24
MYB090 OE	20
MYB090-SRDX OE	4

constituant ainsi une lignée transgénique. Elles ont été isolées et mises en tube dans un milieu favorisant l'enracinement. Un nombre maximal arbitraire de 24 lignées a été isolé pour chaque transformation (tab.III-1). Chaque lignée porte le nom de la transformation (FT OE ou FT-SRDX OE, OE pour over-expressor) suivi d'un numéro.

Les vitro-plants ayant 4 à 6 feuilles (généralement 4 à 5 cm de longueur) sont par la suite acclimatés en serre. L'acclimatation des peupliers transgéniques a été faite en plusieurs vagues (cf. Matériels et Méthodes Tableau 1) pour plusieurs raisons :

- Toutes les transformations n'ont pas été lancées en même temps (conditions expérimentales et matériel du laboratoire optimisé pour traiter 4 transformations différentes à la fois).
- Les micro-tiges transgéniques produites sont isolées au fur et à mesure, classées par taille et multipliées afin d'être maintenues *in vitro* avant l'acclimatation. De plus, elles ne se développent pas toutes à la même vitesse. Il est par conséquent très difficile d'avoir tous les vitro-plants appartenant aux diverses transformations ou même à une seule transformation avec une taille convenable pour l'acclimatation. Dans la mesure du possible, les lignées appartenant à une même expérience de transformation ont été acclimatées en même temps.
- La capacité de la serre destinée à l'accueil exclusif des plantes génétiquement modifiées n'aurait de toutes façons pas permis d'analyser un nombre aussi important de plantes transgéniques simultanément.

Par conséquent, dans les analyses qui suivent, que ce soit les analyses de données de croissance ou les analyses histologiques, chaque lignée transgénique a été comparée au témoin (peupliers sauvages non transformés) de la même vague d'acclimatation afin d'éviter les biais liés aux conditions de culture de chaque vague (photopériode, température, humidité de l'air, incidents éventuels en serre ...). Par ailleurs, pour une meilleure comparaison entre lignée, les données de croissance des lignées transgéniques ont été analysées par rapport aux données de plantes témoins à l'aide d'un ratio (transgénique/témoin).

2- Sélection des lignées d'intérêt

Le choix des lignées intéressantes a été fait en se basant, dans un premier temps, sur le niveau d'expression du transgène (FT ou FT-SRDX) dans le xylème des plantes transgéniques (Annexe 5, criblage niveau 1). Ensuite, l'objectif était de sélectionner, pour les études physiologiques ultérieures, un nombre plus limité de lignées présentant des modifications phénotypiques comparables au sein d'une même transformation que ce soit au niveau histologique et/ou de croissance (Annexe 5, criblage niveau 2). Au final, un maximum de 5 lignées ayant une faible variabilité intra lignée a été sélectionné (Annexe 1, criblage niveau 3). Ainsi, le choix de lignées était comme suit : pour MYB221 OE (n° 1, 8, 11, 12, 18), pour MYB221-SRDX OE (n° 2, 8, 17), pour MYB156 OE (n° 6, 8, 10, 15, 20), pour MYB156-SRDX OE (n° 7, 8, 9, 11), pour MYB090 (n° 6, 9, 15, 19, 23) et pour MYB090-SRDX (n° 1 et 5).

Pour chaque transformation, une lignée ayant un niveau d'expression du transgène faible et un phénotype similaire à celui observé chez le type sauvage a été sélectionnée en plus dans le but de servir de "témoin transgénique" (Annexe 5, tt) afin de lier les phénotypes observés au niveau d'expression du transgène. Pour MYB221 OE, MYB156 OE, MYB090 OE, MYB221-SRDX OE, MYB156-SRDX OE et MYB090-SRDX OE, les lignées "témoin" sont, respectivement, les lignées n° 7, 18, 1, 18, 16 et 3 pour chaque transformation.

3- Caractérisation phénotypique et moléculaire des plantes transgéniques

Lors d'une première expérimentation menée en 2012, nous avons analysé les caractéristiques de croissance des plantes transformées tels que la hauteur de la tige, le diamètre, la vitesse de l'élongation de la tige, la vitesse de croissance radiale (dans certains cas) et le nombre de branches sylleptiques. Nous avons également relevé des observations de phénotypes anormaux au niveau des feuilles, de la tige et d'autres organes de la plante, le cas échéant. Nous avons comparé les lignées transgéniques avec le type sauvage mais aussi, dans certains cas, avec la lignée choisie en tant que "témoin transgénique". Le niveau d'expression du transgène estimé par RT-PCR semi-quantitative est présenté avec les données de croissance dans les figures qui suivent afin d'avoir une idée sur la relation entre le phénotype observé et le niveau d'expression du transgène.

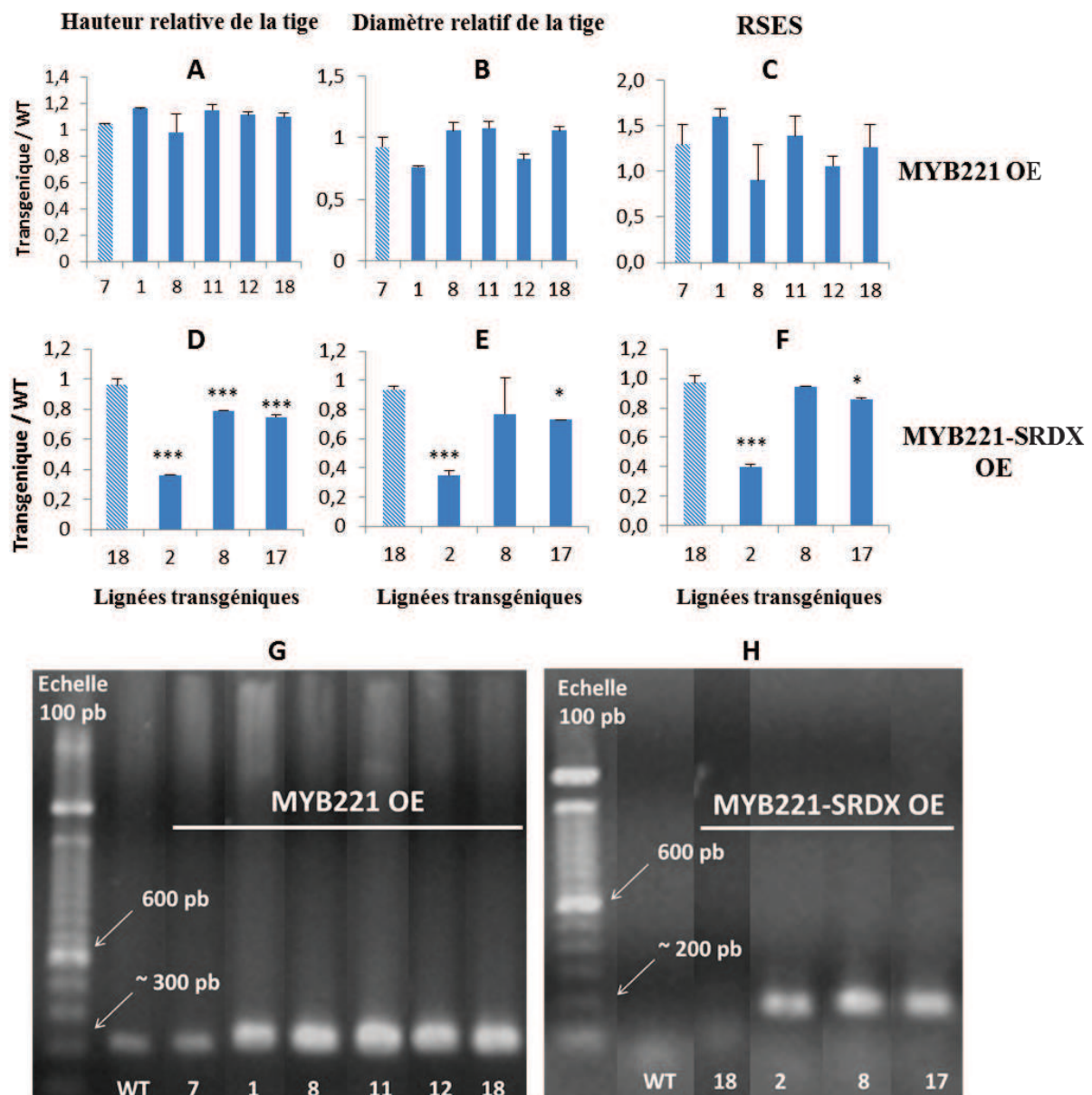


Figure III.1: Caractéristiques de croissance et expression du transgène chez les plantes MYB221 OE et MYB221-SRDX OE.

La hauteur (A, D) et le diamètre (B, E) de tige des peupliers transgéniques MYB221 OE ou MYB221-SRDX OE ont été rapportés, respectivement, à la hauteur et au diamètre de la tige des plantes sauvages (WT). Les valeurs représentées sont ainsi le quotient des valeurs mesurées chez les plantes transformées divisées par les valeurs mesurées chez WT. Les valeurs correspondant au RSES (Vitesse relative d'élongation de la tige) sont calculées à l'aide de cette formule: (Hauteur de tige à 10 semaines – Hauteur de tige à 8 semaines des plantes transgéniques) / (Hauteur de tige à 10 semaines – Hauteur de tige à 8 semaines des plantes sauvages). Les rectangles colorés en bleu représentent les lignées transgéniques sélectionnées. Les rectangles avec un motif rayé représentent les lignées transgéniques considérées comme "témoin transgénique". Le niveau d'expression du transgène a été évalué par RT-PCR semi-quantitative (G, H). Les chiffres en dessous des bandes indiquent les numéros des lignées transgéniques. Les flèches indiquent le poids moléculaire. Les barres d'erreur représentent l'erreur type de la moyenne. Les différences significatives entre les valeurs correspondantes aux plantes transgéniques et celles du type sauvage sont représentées par (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

De plus, pour mieux comprendre le rôle des facteurs de transcription MYB221, MYB156 et MYB090, nous avons analysé des données histologiques provenant de coupes transversales de tige. Pour cela, nous avons réalisé une coloration à la Safranine-Bleu Alcian qui colore en rouge les parois cellulaires lignifiées et en bleu les parois cellulaires riches en cellulose ou en polysaccharides de façon générale.

Par la suite, sur un nombre plus limité de lignées et sur une seconde expérimentation menée en serre en 2013, nous avons quantifié la biomasse aérienne produite et analysé le taux de lignines totales à l'aide de la méthode Klason, chez les plantes transgéniques à l'exception des peupliers MYB156-SRDX OE, faute de matériel végétal (plantes naines et forte mortalité lors de l'acclimatation en serre). Le niveau d'expression du transgène a été également quantifié chez les lignées étudiées afin de vérifier que les mesures observées sont en liaison avec la sur-expression de celui-ci. Nous avons également déterminé l'expression de certains gènes dans le jeune xylème, gènes impliqués dans les voies de biosynthèse des lignines, cellulose et xylanes, et gènes potentiellement cibles de MYB221 ou MYB090 (Chapitre 2).

3.1- Peupliers MYB221 OE et MYB221-SRDX OE

3.1.1- Caractéristiques de croissance

Les plantes MYB221 OE (lignées 1, 8, 11, 12 et 18) n'ont pas montré de différence significative avec le type sauvage ni avec le 'témoin transgénique' (lignée 7) en ce qui concerne la hauteur, le diamètre et la vitesse d'élongation de la tige (figIII.1-A, B, C). Au contraire, les peupliers MYB221-SRDX OE et surtout la lignée n°2 ont clairement une tige plus courte, de diamètre plus faible et une vitesse d'élongation de tige moins importante en comparaison avec les témoins (figIII.1-D, E, F), à l'exception de la lignée 8 qui a un diamètre de tige et un taux de croissance en hauteur comparable au type sauvage et au témoin transgénique (lignée 18) (figIII.1-E, F).

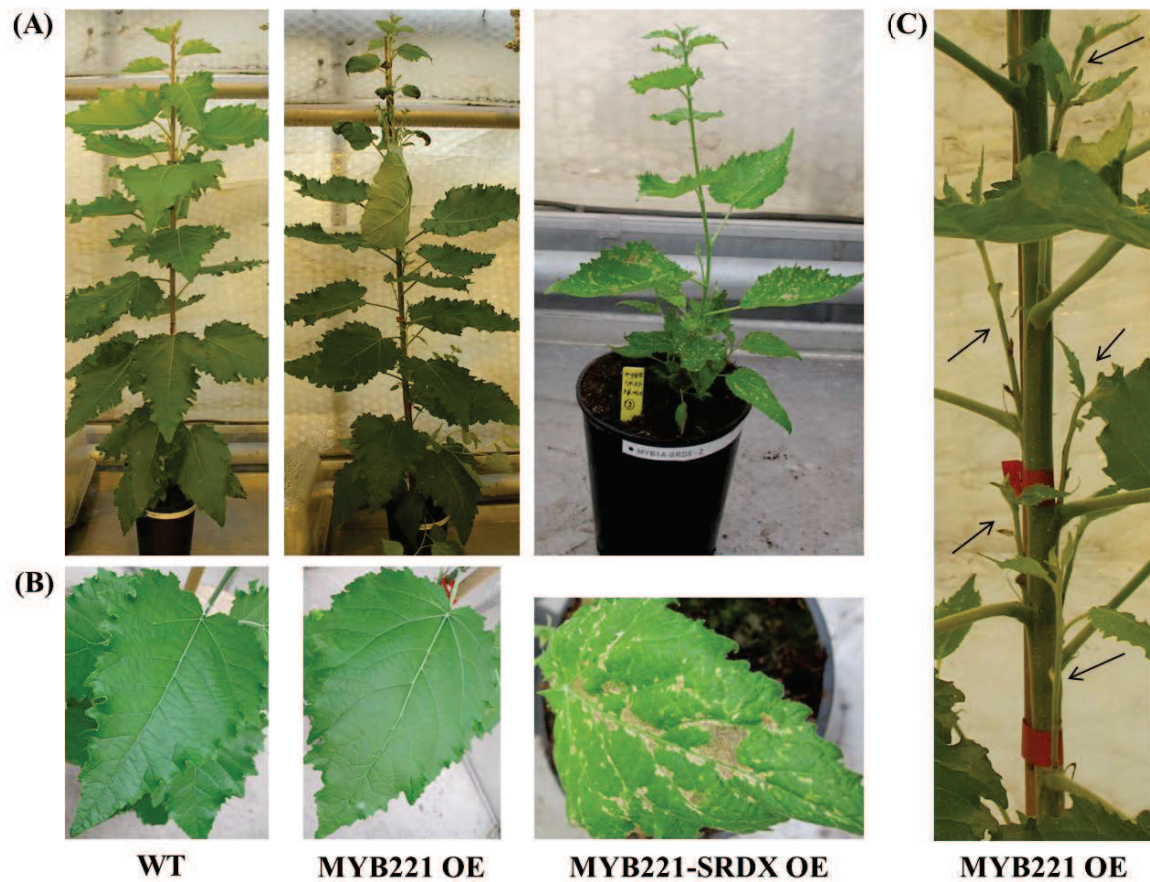


Figure III.3: Comparaison entre le phénotype des plantes MYB221 OE, MYB221-SRDX OE et les plantes sauvages.

Le phénotype des plantes MYB221 OE et MYB221-SRDX OE a été comparé avec celui des plantes sauvages (WT) au niveau de la plante entière (A) et des feuilles (B). Les branches sylleptiques des plantes MYB221 OE (C) sont indiquées par des flèches.

Le nombre de branches sylleptiques a augmenté d'une façon spectaculaire chez les plants MYB221 OE en comparaison avec le type sauvage et la lignée 7 (figIII.2 ; 3-A, C). Il est situé dans un intervalle entre, environ 10 (lignée 1 et 8) et 20 (lignée 18) branches alors que le sauvage et le "témoin transgénique" ont une moyenne inférieure à 1 branche. Par contre, les lignées MYB221-SRDX OE n° 2, 8 et 17 ne présentent pas de différence significative avec le témoin, pour ce caractère (figIII.2). Au niveau foliaire, les plants MYB221-SRDX OE ont développé des feuilles déformées et plus petites (figIII.3-B) et présentent une difficulté plus ou moins importante à se tenir droits sans tuteurage (figIII.3-A). Ceci n'est pas le cas pour les peupliers MYB221 OE qui ont des feuilles similaires à celles du sauvage ainsi qu'un port normal (figIII.3-A, B).

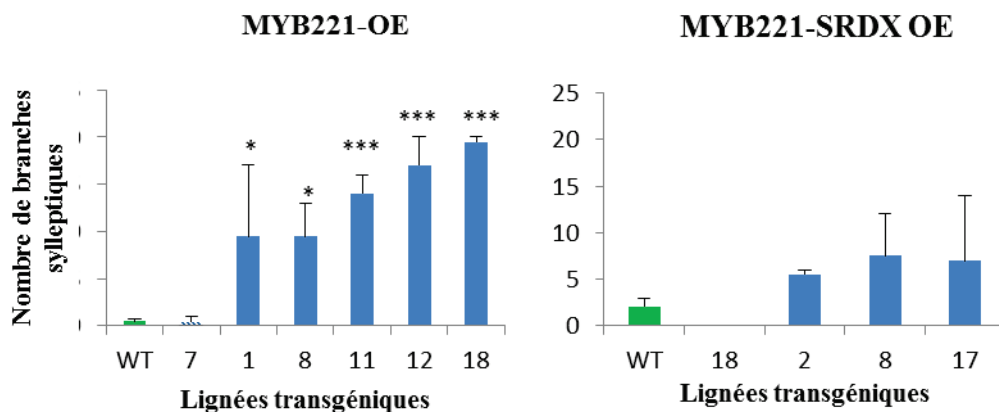


Figure III.2: Nombre de branches sylleptiques chez les plantes MYB221 OE et MYB221-SRDX OE.

Le nombre de branches sylleptiques a été compté après 10 semaines de croissance. Les rectangles colorés en bleu représentent les lignées transgéniques sélectionnées. Les rectangles avec un motif rayé représentent les lignées transgéniques considérées comme "témoins transgéniques". Les rectangles colorés en vert illustrent les valeurs des plantes sauvages (WT). Les barres d'erreur représentent l'erreur type de la moyenne de 2 échantillons. Les différences significatives entre les valeurs correspondantes aux plantes transgéniques et celles du type sauvage sont représentées par (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

3.1.2- Caractéristiques histologiques

Les plantes MYB221 OE n'ont pas montré de différences notables au niveau histologique avec les plantes sauvages. En effet, dans les sections colorées, que ce soit des peupliers transgéniques ou des témoins, on note la même organisation cellulaire, en terme de rayons, fibres et vaisseaux (nombre, diamètre). Les parois cellulaires des fibres de xylème, des éléments de vaisseaux et des rayons parenchymateux sont colorées en rouge de la même façon que le

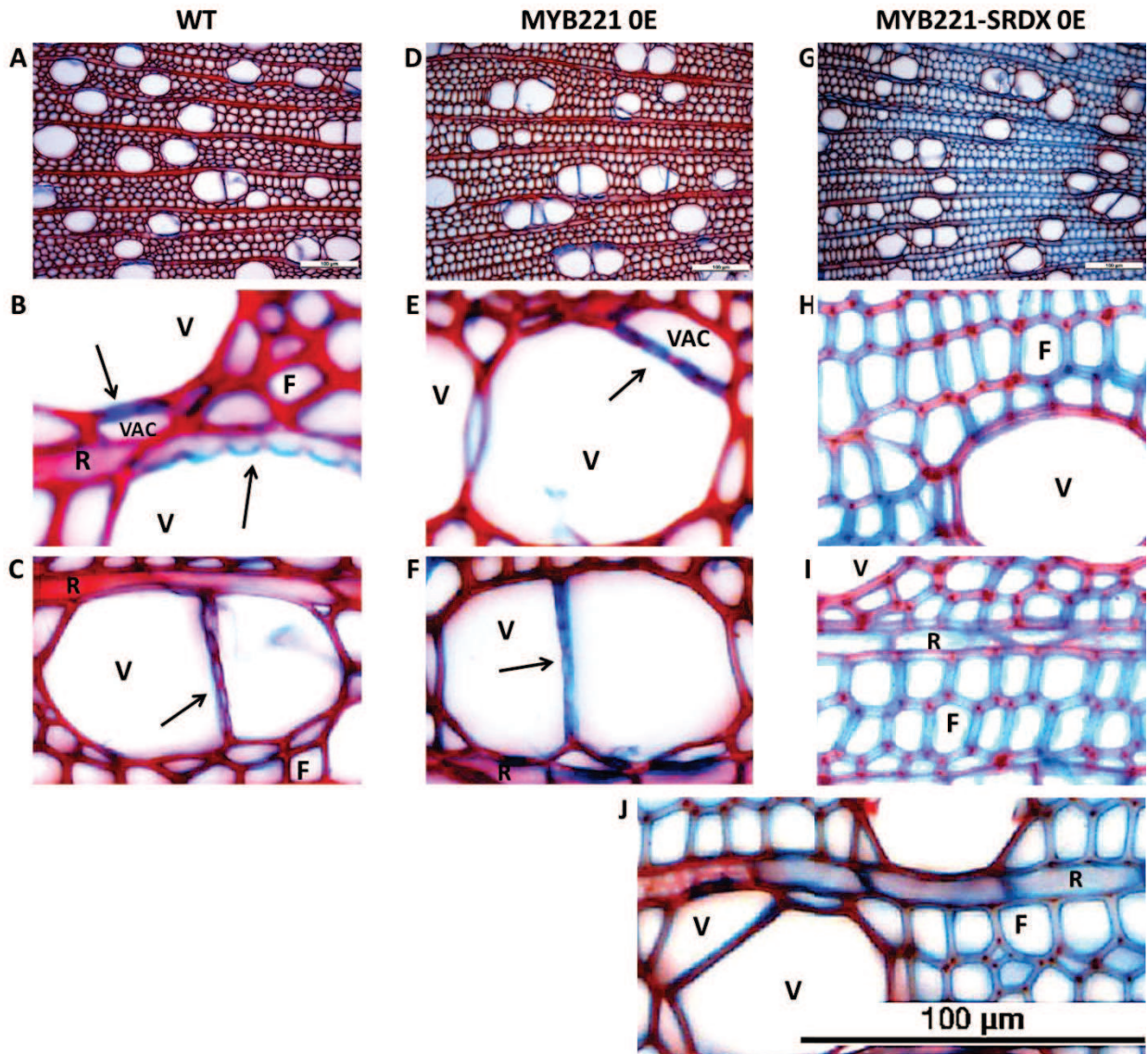


Figure III.4 : Analyse histologique des plantes MYB221 OE et MYB221-SRDX OE.

Les coupes transversales de tiges de peupliers sauvages non transformés génétiquement *Populus tremula x alba* clone INRA 717-1B4 (WT) (A, B, C), de plantes MYB221 OE (D, E, F) et MYB221-SRDX OE (G, H, I, J) ont été colorées par la Safranine-Bleu Alcian. Les flèches indiquent les ponctuations. Les photos B&C, E&F, H&I&J sont des agrandissements issus des photos A, D et G respectivement. V, éléments de vaisseaux. F, fibres de xylème. R, rayons ligneux. VAC, cellules associées aux vaisseaux. Barre d'échelle = 100 µm.

témoin (figIII.4-A, D). De plus, on peut observer des ponctuations de moins de 5 μm de diamètre, plus ou moins circulaires en contact avec les cellules voisines, situées sur les côtés des parois des éléments de vaisseaux. Ces structures sont localisées entre deux éléments de vaisseau, entre une cellule de rayon et un élément de vaisseau, ou bien entre une cellule associée aux vaisseaux et un élément de vaisseau. Elles sont colorées en bleu signifiant qu'elles ne sont pas lignifiées, que ce soit dans les plantes transgéniques ou les plantes témoins (figIII.4-B, C, E, F).

En revanche, les coupes transversales des peupliers MYB221-SRDX OE ont montré une modification importante de la lignification des fibres et des rayons au moins au niveau de plusieurs plages cellulaires de la coupe. Les parois des éléments des vaisseaux présentent toujours une couleur rouge caractéristique. Par contre, les parois cellulaires des fibres dans ces plages sont devenues bleu clair indiquant une perturbation de leur lignification. Seuls les coins de certaines cellules de fibres restent lignifiés à l'exception des parois d'autres fibres qui sont légèrement lignifiées mais de façon beaucoup moins importante que chez le sauvage (figIII.4-G, H). De plus, la paroi de certaines cellules de parenchyme dans les zones modifiées est également aussi colorée en bleu clair et apparaît donc moins lignifiée (figIII.4-G, I); celles en contact avec les éléments de vaisseaux ont une lignification normale (figIII.4-G, J).

On note qu'il n'était pas techniquement possible d'analyser toutes les lignées transgéniques en utilisant la coloration à la Safranine-Bleu Alcian car cette technique n'est pas adaptée aux coupes de tiges ayant un diamètre très petit. Ainsi, les coupes de tiges de plants MYB221-SRDX OE ayant une tige minuscule ont été analysées à l'aide du microscope confocal capable de révéler l'autofluorescence des lignines au niveau des parois cellulaires. Les images obtenues ont montré une énorme baisse de l'autofluorescence des lignines chez les peupliers transformés en comparaison aux peupliers sauvages. Le niveau d'autofluorescence est faible au niveau des parois des fibres, des rayons mais aussi des éléments de vaisseaux (figIII.5) suggérant une baisse plus forte de la lignification chez les lignées ayant une tige très petite.

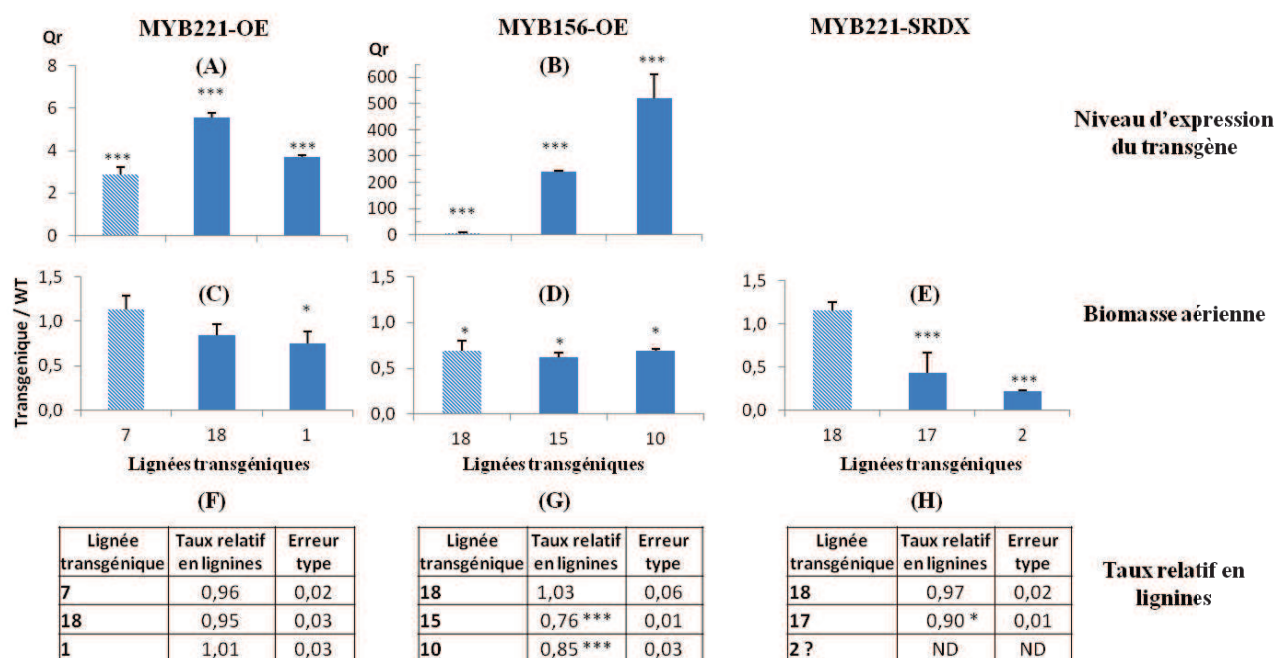


Figure III.6: Mesure du taux de lignines totales, de biomasse et d'expression du transgène chez les peupliers MYB221 OE, MYB156 OE et/ou MYB221-SRDX OE.

Le taux de lignines totales a été mesuré par la méthode de Klason (F, G, H). Toute la partie aérienne des plantes (tige + feuilles) a été pesée pour évaluer la biomasse aérienne. Le taux de lignines totales et la biomasse aérienne ont été rapportés aux valeurs mesurées chez les plantes sauvages. Qr représente l'expression quantitative de *MYB221* (A) et *MYB156* (B) rapportée à celle mesurée chez le sauvage. Les rectangles pleins colorés en bleu représentent les lignées transgéniques d'intérêt. Les rectangles avec un motif rayé représentent les "témoins transgéniques". Les barres d'erreur représentent l'erreur type de la moyenne. Les différences significatives entre chaque lignée transgénique et le témoin ont été représentées par : (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

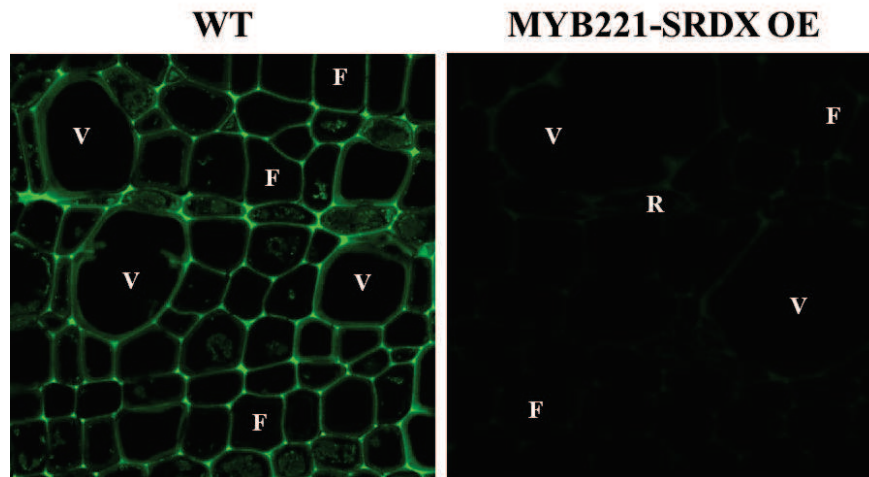


Figure III.5: Auto-fluorescence des lignines chez les plantes MYB221-SRDX OE. L'autofluorescence de lignines a été observée au niveau de coupes transversales de tige de peupliers sauvages *Populus tremula* x *alba* clone INRA 717-1B4 (WT) et des plantes transgéniques MYB221-SRDX OE (lignée 2) à l'aide d'un microscope confocal. Agrandissement: x40. V, éléments de vaisseaux, F, fibres de xylème. R, rayons ligneux.

3.1.3- Analyses biochimiques et de biomasse

Ces expériences ont été réalisées sur une nouvelle série de plantes acclimatées en même temps en 2013 et sur un nombre plus restreint de lignées transgéniques (2 à 3).

Les peupliers MYB221 OE n'ont pas montré de variation du taux de lignines ni de biomasse par rapport aux plantes témoins (sauvages et transgéniques) à l'exception de la lignée 1 qui a moins de biomasse. Ceci peut être expliqué par le fait que cette lignée possède un diamètre de tige plus petit que celui des plantes sauvages (figIII.1-B). L'expression de *MYB221* chez les plantes transgéniques est relativement peu élevée bien que le transgène soit sous le contrôle du promoteur fort CaMV35S doublé : les lignées 1 et 18 ont respectivement environ 3,5 et 5 fois plus de transcrits que le sauvage. Quant à la lignée 7, elle a 2,5 à 3 fois plus de transcrits que le sauvage (figIII.6-A).

Chez les peupliers MYB221-SRDX OE, le taux de lignines devrait être plus faible par rapport aux plantes témoins (type sauvage et 'témoin transgénique') si on se base sur les données histologiques. La figure III.6-H montre que la lignée 17 possède environ 10 % de lignines en moins que les plantes sauvages. Cependant, nous avons rencontré un problème technique empêchant malheureusement de mesurer le taux de lignines des plantes de la lignée 2 (pas assez de poudre de tiges broyées, due à la petite taille de ces plantes). Cette quantité n'était pas suffisante pour avoir une mesure fiable du taux de lignines. La figure III.6-E illustre

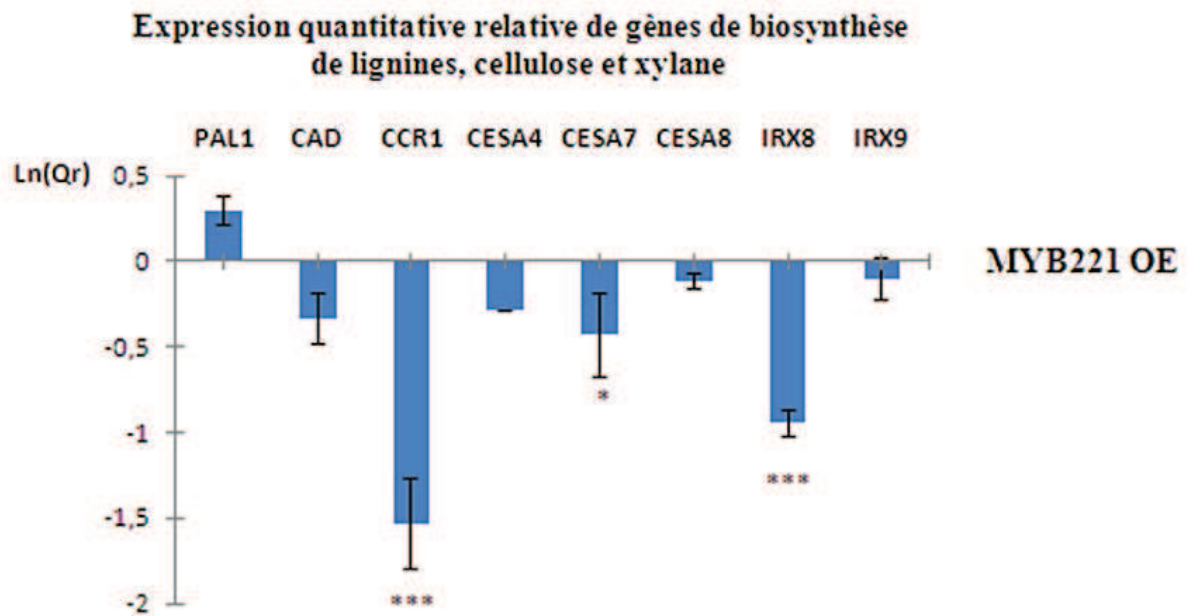


Figure III.7: Régulation de gènes de biosynthèse de lignines, cellulose et xylanes dans le xylème des plantes MYB221 OE.

Le logarithme des valeurs de l'expression quantitative relative ($\text{Ln}(\text{Qr})$) des gènes de biosynthèse de lignines (*PAL1*, *CAD*, *CCR1*), cellulose (*CESA4*, *CESA7*, *CESA8*) et xylanes (*IRX8*, *IRX9*) au niveau du xylème de plantes transgéniques sur-exprimant *MYB221* est présenté par des rectangles bleus. Les barres d'erreur représentent l'erreur-type de la moyenne. La différence significative entre les valeurs d'expression de plantes transgéniques et celles de plantes sauvages est représentée par (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

une baisse importante de la biomasse des lignées transgéniques. Cette baisse est d'environ 50 % et 80 % respectivement, chez les lignées 17 et 2 par rapport aux plantes sauvages et à la lignée témoin. Ceci est en parfait accord avec les données de croissance de ces plantes.

3.1.4- Expression de gènes de biosynthèse des lignines, cellulose et xylanes

Nous avons suivi l'expression de gènes majeurs impliqués dans ces différentes voies métaboliques. Nous avons tout d'abord choisi 3 gènes situés à des niveaux différents de la voie de biosynthèse des lignines ; *PAL1* (Potri.006G126800) qui intervient à une étape précoce de cette voie, *CCR1* (Potri.003G181400) et *CAD* (Potri.009G095800) qui jouent un rôle déterminant à un stade plus avancé de cette voie. De plus, nous avons également étudié l'expression de gènes de biosynthèse de cellulose (*CESA4* : Potri.002G257900, *CESA7* : Potri.006G181900, *CESA8* : Potri.011G069600) et de xylanes (*IRX8* : Potri.011G132600 et *IRX9* : Potri.016G086400) chez les plantes transformées.

La PCR quantitative relative a montré que, chez les plantes MYB221 OE, l'expression du gène *CCR1* est fortement réprimée, alors que l'expression de *PAL1* et *CAD* est inchangée (figIII.7). Chez les plantes MYB221 OE, *CESA7* est le seul gène de cellulose synthase à être régulé négativement dans le xylème de ces plants (figIII.7). De même, l'expression du gène *IRX8* est réprimée chez ces transgéniques alors que celle de *IRX9* ne l'est pas (figIII.7).

3.1.5- Expression de gènes cibles identifiés par ChIP-SEQ

Le niveau d'expression d'un nombre de gènes cibles de MYB221 identifiés par la technique ChIP-SEQ a été mesuré chez les plantes transgéniques surexprimant le FT en question ou sa forme de répression dominante dans le but d'avoir une idée sur sa nature régulatrice (répresseur ou activateur ou les deux à la fois).

Les résultats montrent que l'expression de deux gènes cibles (Potri.013G000600 : membrane bound O-acyl transférase (MBOAT) family protein) et (Potri.T011400 : Transcription factor, MYB superfamily) est réprimée chez les plantes MYB221 OE mais pas chez les peupliers MYB221-SRDX OE (figIII.8). L'expression de (Potri.004G067000 : Alcohol déshydrogenase, class III) est par contre activée, mais seulement chez les transgéniques

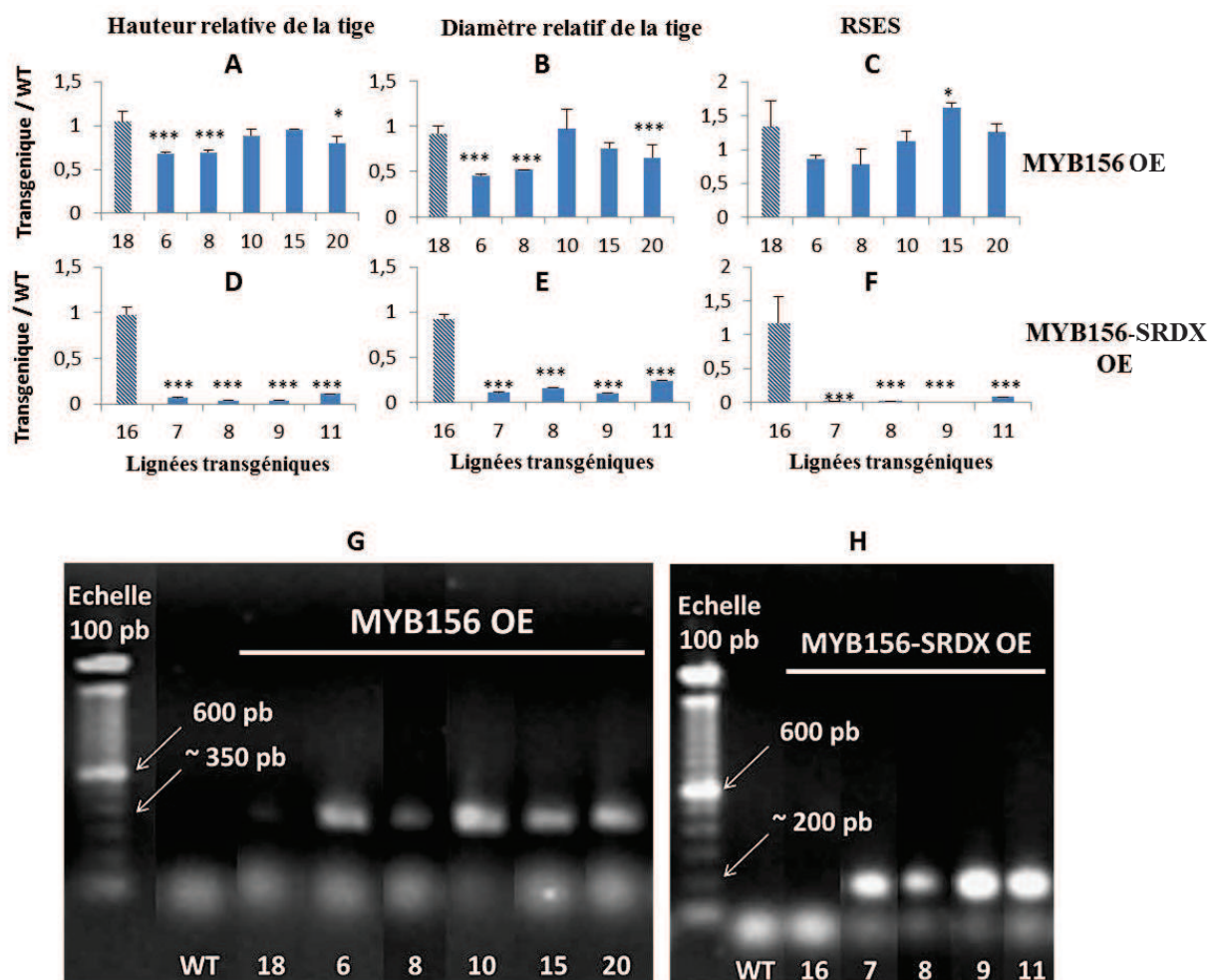


Figure III.9: Caractéristiques de croissance et expression du transgène chez les plantes MYB156 OE et MYB156-SRDX OE.

La hauteur (A, D) et le diamètre (B, E) de tige des peupliers transgéniques MYB156 OE ou MYB156-SRDX OE ont été rapportés, respectivement, à la hauteur et au diamètre de la tige des plantes sauvages (WT). Les valeurs représentées sont ainsi le quotient des valeurs mesurées chez les plantes transformées divisées par les valeurs mesurées chez WT. Les valeurs correspondant au RSES (Vitesse relative de la croissance de la tige) sont calculées à l'aide de cette formule: $(\text{Hauteur de tige à 10 semaines} - \text{Hauteur de tige à 8 semaines des plantes transgéniques}) / (\text{Hauteur de tige à 10 semaines} - \text{Hauteur de tige à 8 semaines des plantes sauvages})$. Les rectangles colorés en bleu représentent les lignées transgéniques sélectionnées. Les rectangles avec un motif rayé représentent les lignées transgéniques considérées comme "témoins transgéniques". Le niveau d'expression du transgène a été évalué par RT-PCR semi-quantitative (G, H). Les chiffres en dessous des bandes indiquent les numéros des lignées transgéniques. Les flèches indiquent le poids moléculaire. Les barres d'erreur représentent l'erreur type de la moyenne. Les différences significatives entre les valeurs correspondantes aux plantes transgéniques et celles du type sauvage sont représentés par (* t-test $P < 0.05$) or (***) t-test $P < 0.01$).

MYB221 OE. L'expression d'une autre cible testée (Potri.006G251800 : bZIP transcription factor) ne change pas entre les lignées issues des deux types de transformation et le sauvage.

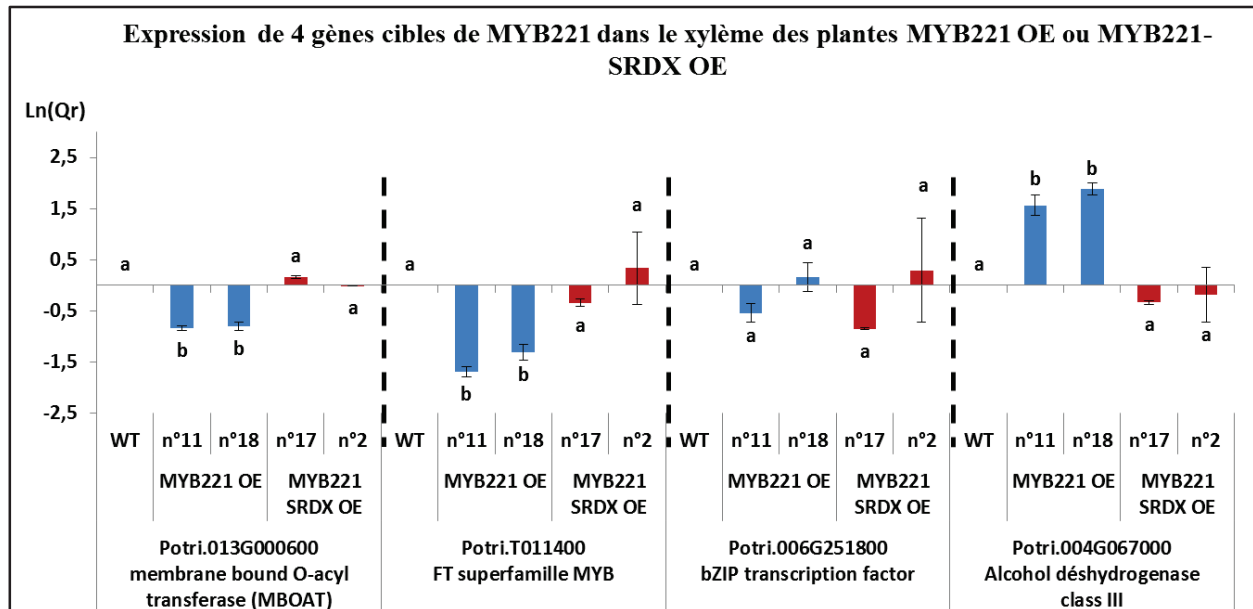


Figure III.8: Expression de 4 gènes cibles de MYB221 dans le xylème des plantes MYB221 OE et MYB221-SRDX OE.

Le logarithme des valeurs de l'expression quantitative relative (Ln(Qr)) de 4 gènes cible de MYB221 identifiés par ChIP-SEQ : Potri.013G000600, Potri.T011400, Potri.006G251800 et Potri.004G067000 au niveau du xylème de plantes transgéniques surexprimant *MYB221* (bleu) et *MYB221-SRDX* (rouge). Les barres d'erreur représentent l'erreur-type de la moyenne. La différence significative du niveau d'expression (Ln(Qr)) entre les différentes lignées y compris les plantes sauvages (WT) est représentée par des lettres différentes (t-test $P < 0.01$) pour chaque gène cible séparément.

3.2- Peupliers MYB156 OE et MYB156-SRDX OE

3.2.1- Caractéristiques de croissance

Les caractéristiques de croissance ont été analysées chez les plantes MYB156 OE. Bien que leur vitesse de croissance en hauteur entre 8 et 10 semaines, à l'exception de la lignée 15, est invariable par rapport aux plantes sauvages et au 'témoin transgénique' (lignée 18), la hauteur et le diamètre de la tige de plusieurs lignées (6, 8, 20) sont plus faibles (figIII.9-A, B, C). Les deux lignées 10 et 15 n'ont pas montré de différence avec les témoins (figIII.9-A, B, C) concernant ces deux mesures. Nous avons également procédé à l'analyse de la vitesse de l'élongation de la tige (RSES) et de la croissance radiale (RRGS) au niveau de 3 intervalles différents de la période de croissance chez des plantes de la lignée 15. Ces mesures montrent

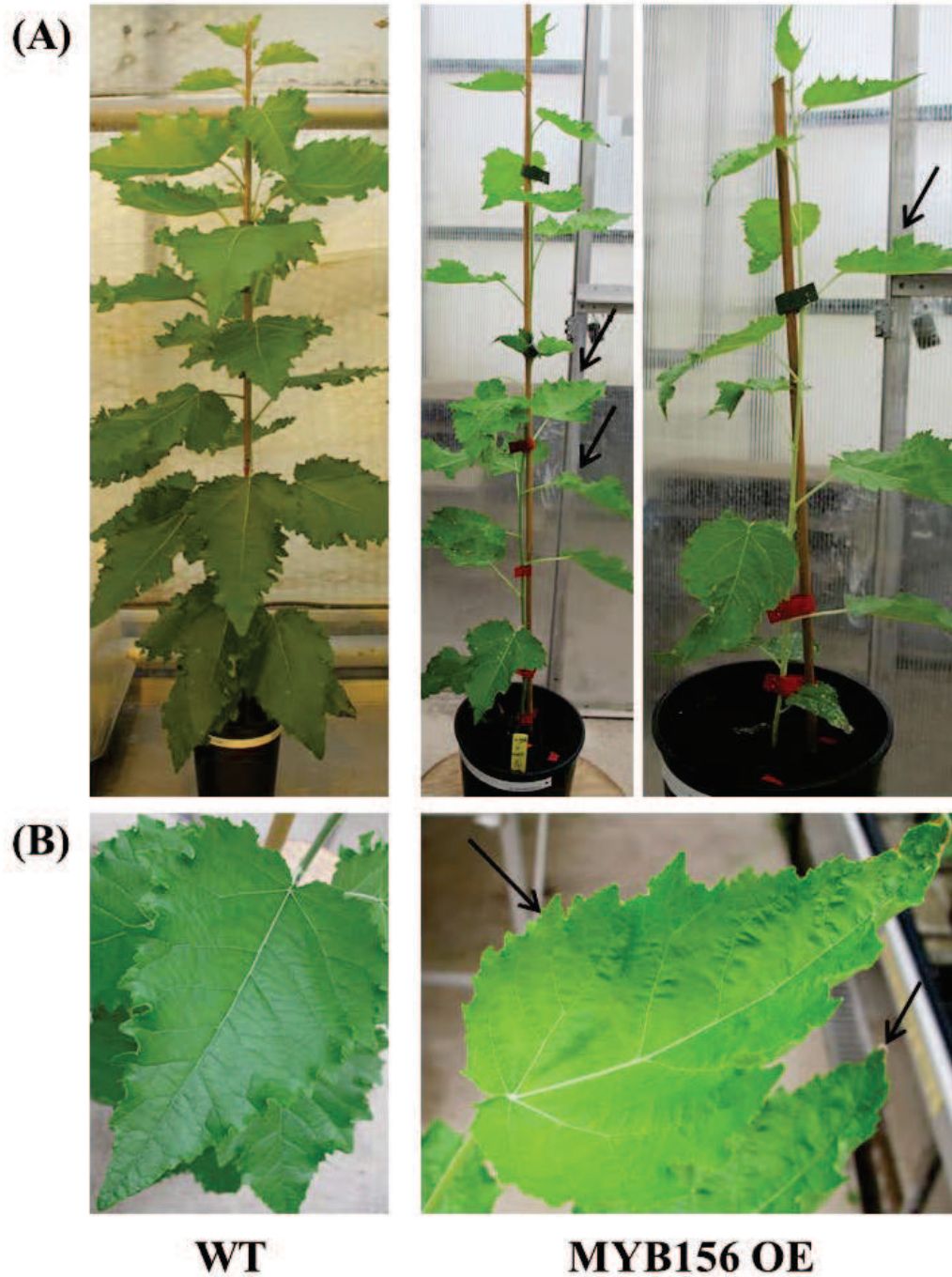


Figure III.11: Comparaison entre le phénotype des plantes MYB156 OE et les plantes sauvages.

Le phénotype des plantes MYB156 OE (lignée 4 et 1 de gauche vers la droite) a été comparé avec celui des plantes sauvages (WT) au niveau de la plante entière (A) et des feuilles (B). Les flèches indiquent les feuilles en forme de cuvette de plantes MYB156 OE.

une baisse d'environ 50 % de la vitesse de croissance des tiges de ces plants par rapport aux plantes sauvages pendant les 5 premières semaines de croissance. Ensuite, cette vitesse a augmenté de façon importante devenant même supérieure à celle de type sauvage en ce qui concerne la croissance en hauteur (figIII.10). Les plantes MYB156 OE ont montré également une certaine difficulté, plus ou moins importante selon les lignées, pour se tenir droites sans tuteurage d'où l'observation de déformations au niveau de la tige. De plus, ces plantes possèdent des feuilles généralement de taille plus ou moins petite (selon les lignées) et légèrement déformées (plus concaves / en forme de cuvette) (figIII.11-A, B).

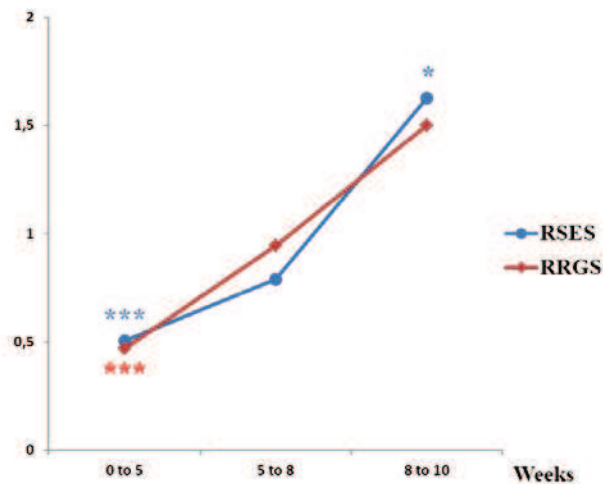


Figure III.10: Evolution de la vitesse de croissance de tige des plantes MYB156 OE (lignée 15).

La vitesse relative de la croissance en hauteur (RSES)(cercles bleus pleins) et la vitesse relative de croissance radiale (RRGS) (losanges rouges pleins) de la tige ont été calculées en divisant les valeurs mesurées chez les plantes transgéniques MYB156 OE (lignée 15) sur les valeurs mesurées chez les plantes sauvages au niveau de 3 intervalles de croissance: jusqu'à 5 semaines (0 to 5), de la 5^{ème} semaine à la 8^{ème} semaine (5 to 8) et de la 8^{ème} semaine à la 10^{ème} semaine (8 to 10). Les différences significatives entre les valeurs correspondantes aux plantes transgéniques et celles du type sauvage sont représentées par (* t-test $P < 0.05$) or (***) t-test $P < 0.01$).

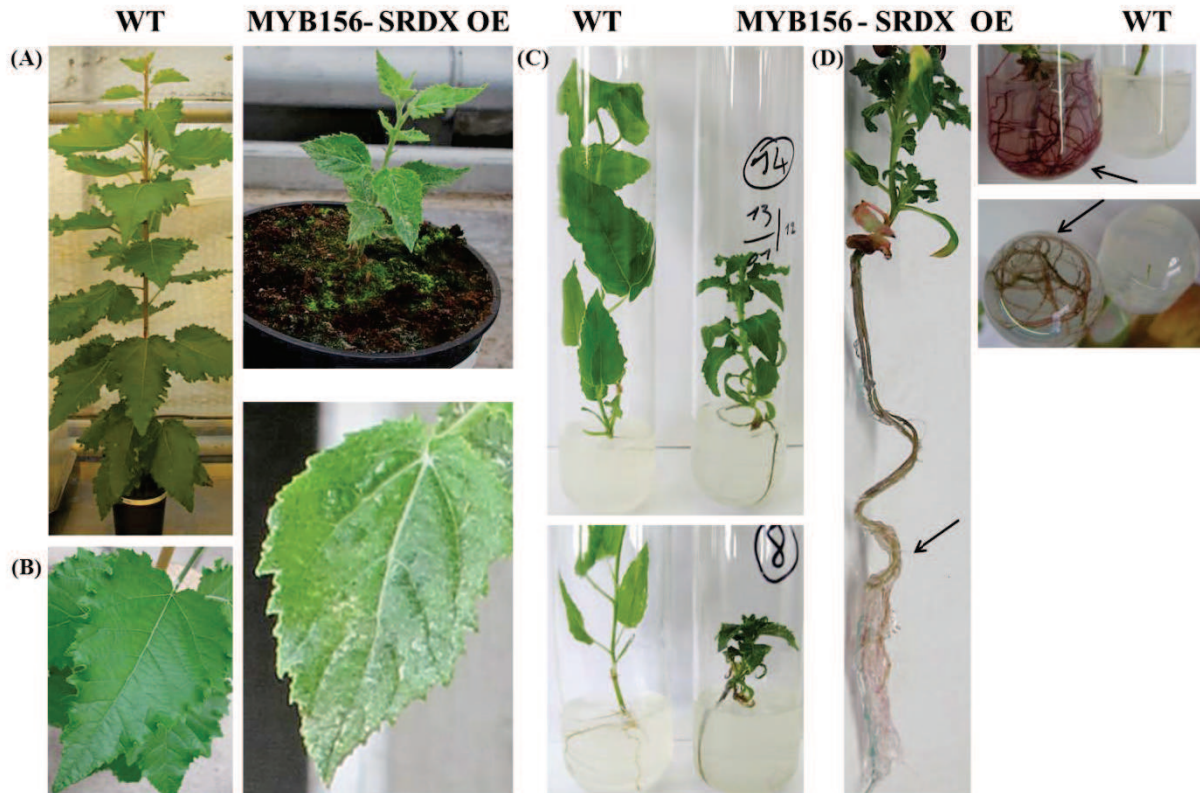


Figure III.12: Comparaison du phénotype des plantes MYB156-SRDX OE et les plantes sauvages (WT).

Le phénotype en serre des plantes MYB156-SRDX OE (lignée 22) a été comparé avec celui des plantes sauvages (WT) au niveau de la plante entière (A) et des feuilles (B). Le phénotype *in vitro* des plantes MYB156-SRDX OE (lignées 14 et 8) a été comparé avec celui des plantes sauvages (WT) au niveau de la plante entière (C) et des racines (D). Les flèches indiquent le système racinaire très développé des plantes MYB156-SRDX OE.

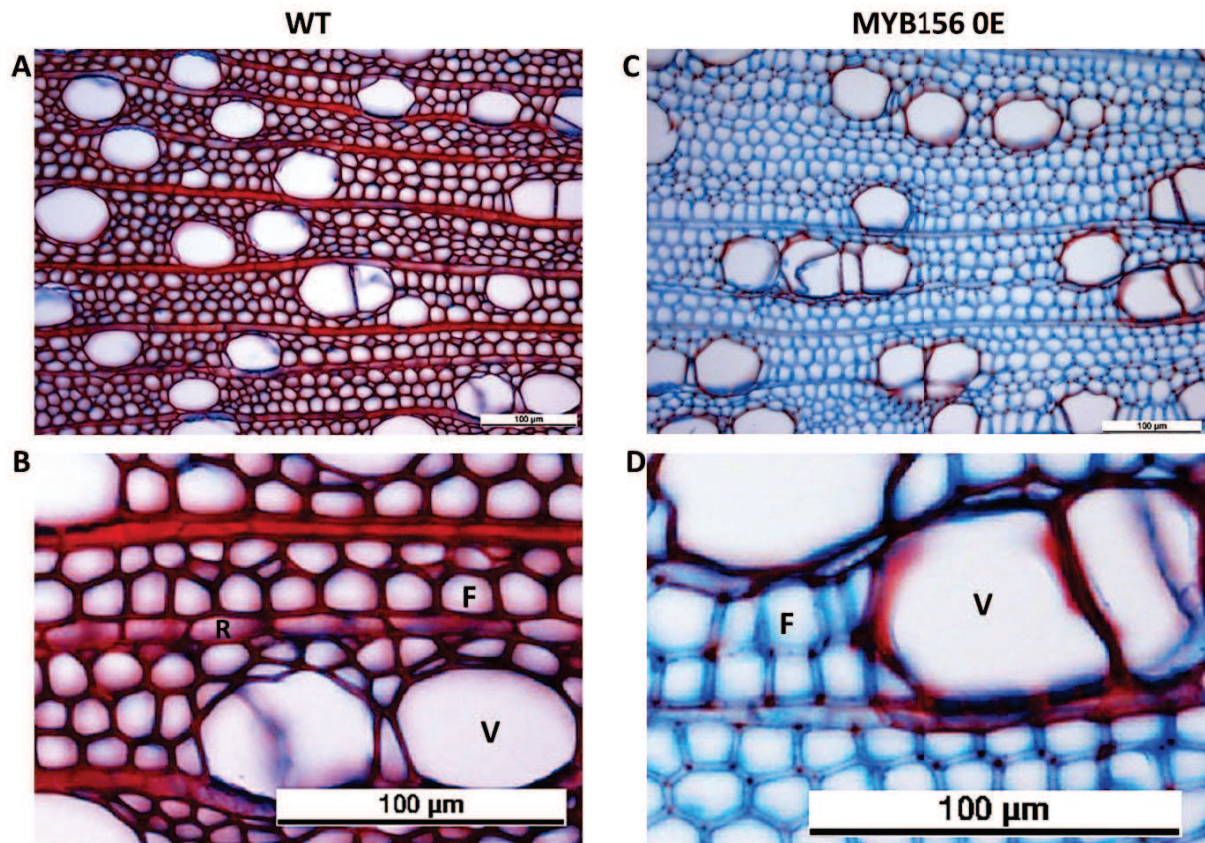


Figure III.13: Analyse histologique des plantes MYB156 OE.

Les coupes transversales de tiges de peupliers sauvages non transformés génétiquement *Populus tremula x alba* clone INRA 717-1B4 (WT) (A, B) et des plantes MYB156 OE (lignée 15) (C, D) ont été colorées par la Safranine-Bleu Alcian. Les photos B et D sont des agrandissements issus des photos A et C respectivement. V, éléments de vaisseaux. F, fibres de xylème. R, rayons ligneux. Barre d'échelle = 100 μm.

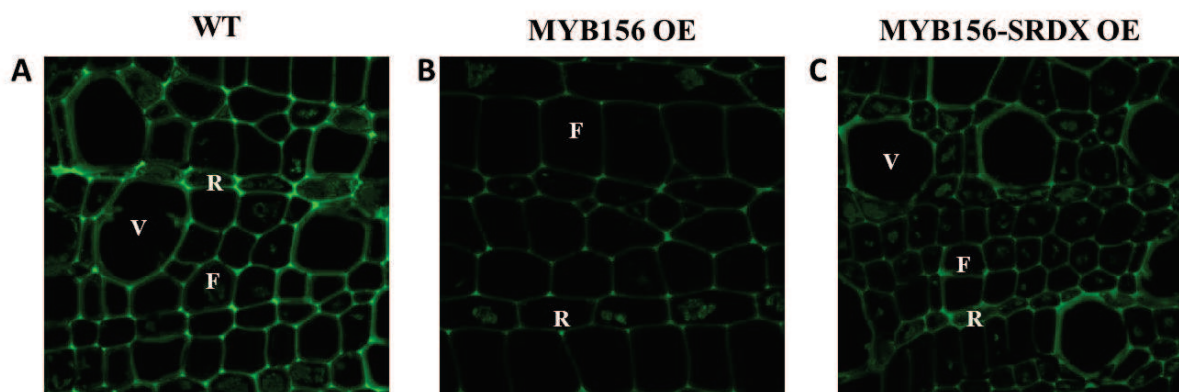


Figure III.14: Auto-fluorescence des lignines chez les plantes MYB156 OE et MYB156-SRDX OE.

L'auto-fluorescence de lignines a été observée au niveau de coupes transversales de tige de peupliers sauvages *Populus tremula x alba* clone INRA 717-1B4 (WT) et des plantes transgéniques MYB156 OE (lignée 22) et MYB156-SRDX OE (lignée 9) à l'aide d'un microscope confocal. Agrandissement : x40. V, éléments de vaisseaux, F, fibres de xylème. R, rayons ligneux.

La croissance des peupliers MYB156-SRDX OE est très fortement impactée puisque les plantes transformées sont toutes naines. En effet, elles ont une tige très courte avec un diamètre minuscule et une très faible vitesse d'élongation (figIII.9-D, E, F). Ces plantes se caractérisent aussi par des feuilles de taille très réduite et des entrenœuds très courts (figIII.12-A, B). Ce phénotype est facilement identifié en conditions de culture *in vitro* (figIII.12-C). On peut remarquer aussi que ces plantes ont des racines développées de telle façon qu'elles étaient enroulées à la base du tube de culture (figIII.12-D). L'acclimatation de ces plants nains en serre s'est avérée très difficile à cause d'une très forte mortalité après le transfert en serre.

3.2.2- Caractéristiques histologiques

De même que pour les plants MYB221-SRDX OE, la coloration à la Safranine-Bleu Alcian a montré une forte modification de la lignification des parois cellulaires des fibres des plantes MYB156 OE (figIII.13-A, C) à l'exception des coins de certaines d'entre elles qui sont légèrement lignifiés (figIII.13-B, C). Les cellules des rayons ligneux ont aussi des parois moins lignifiées chez les transgéniques par comparaison au type sauvage sauf certaines cellules de parenchyme en contact avec les éléments des vaisseaux qui ont des parois lignifiées (figIII.13-D). Les plantes de très petite taille ont été analysées également au moyen de l'auto-fluorescence des lignines. Les images obtenues montrent que les parois des fibres des plantes transformées sont moins lignifiées que celles des plantes sauvages à l'exception des coins de certaines parois qui sont lignifiées ponctuellement (figIII.14-B).

Chez les plants MYB156-SRDX OE, la lignification des parois des fibres semble également modifiée par rapport à celle du témoin. L'auto-fluorescence des lignines au niveau de la paroi des éléments des vaisseaux apparaît toutefois inchangée entre les plantes transgéniques et les plantes sauvages (figIII.14-C).

3.2.3- Analyses biochimiques et de biomasse

Les lignées 15 et 10 des plants MYB156 OE ont montré une forte expression de *MYB156*. Ce niveau d'expression est de loin supérieur à celui observé chez la lignée 18, considérée comme un témoin transgénique, et pour laquelle le transgène est 7 fois plus exprimé que chez le sauvage (figIII.6-B). Alors que le taux de lignines est inchangé chez la lignée 18 (figIII.6-G) qui a un faible niveau d'expression du transgène (figIII.6-B), on enregistre une

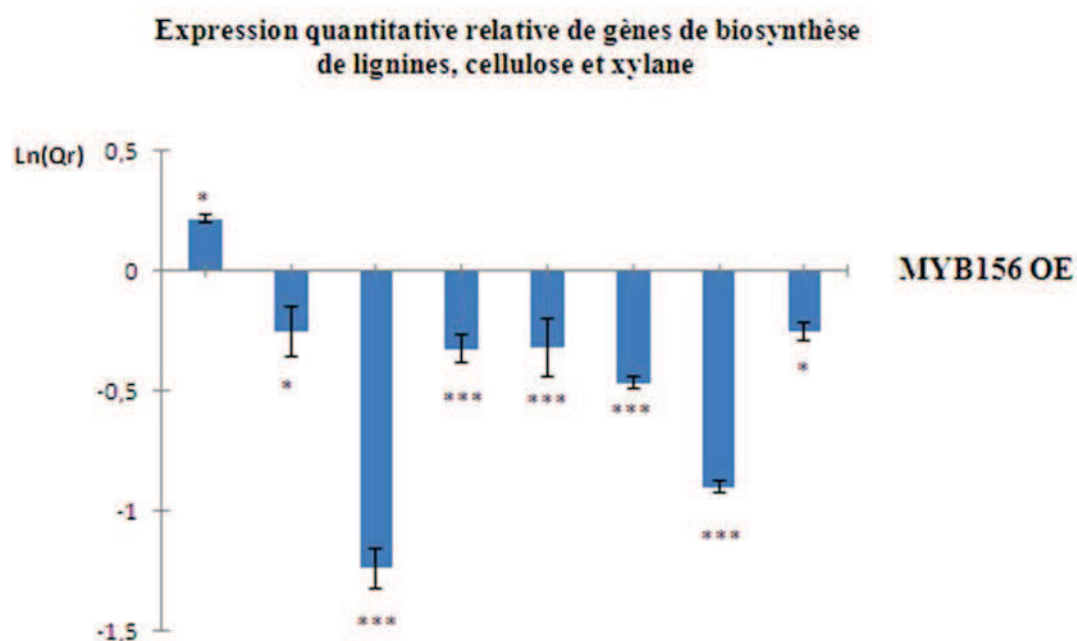


Figure III.15: Régulation de gènes de biosynthèse de lignines, cellulose et xylanes dans le xylème des plantes MYB156 OE.

Le logarithme des valeurs de l'expression quantitative relative ($\text{Ln}(\text{Qr})$) des gènes de biosynthèse des lignines (*PAL1*, *CAD*, *CCR1*), cellulose (*CESA4*, *CESA7*, *CESA8*) et xylanes (*IRX8*, *IRX9*) au niveau du xylème de plantes transgéniques sur-exprimant *MYB156* est présenté par des rectangles bleus. Les barres d'erreur représentent l'erreur-type de la moyenne. La différence significative entre les valeurs d'expression de plantes transgéniques et celles de plantes sauvages est représentée par (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

baisse de 15% et 25% chez les lignées 10 et 15 respectivement. D'un autre côté, nous n'avons pas observé de relation claire entre le niveau d'expression du transgène et le taux de lignines d'une part, et la masse de la tige d'autre part, car toutes les lignées transgéniques étudiées, y compris le 'témoin transgénique' ont un niveau de biomasse significativement plus bas que celui du type sauvage (figIII.6-D).

3.2.4- Expression de gènes de biosynthèse des lignines, cellulose et xylanes

Les résultats ont montré que l'expression du gène *CCR1* est réprimée ainsi que celle du gène *CAD*, mais de façon plus légère. En revanche, l'expression de *PAL1* est régulée positivement chez ces derniers (figIII.15). D'un autre côté, nous avons trouvé que l'expression de tous les gènes de biosynthèse de cellulose testés est réprimée chez les plantes MYB156 OE (figIII.15). De même, les gènes *IRX8* et *IRX9* sont réprimés chez ces plantes (figIII.15).

3.3- Peupliers MYB090 OE et MYB090-SRDX OE

Concernant les plants MYB090 et MYB090-SRDX, nous n'avons pas pu étudier en parallèle des témoins transgéniques car un nombre de plants appartenant à cette catégorie a été perdu à cause d'un incident imprévu (problème d'arrosage ponctuel) empêchant ainsi la réalisation des analyses statistiques sur ces témoins.

3.3.1- Caractéristiques de croissance

Nous avons tout d'abord analysé la hauteur et le diamètre des tiges de MYB090 OE et MYB090-SRDX OE. Nous observons une baisse de la hauteur de la tige par comparaison avec les peupliers non transformés. La diminution de la taille des plantes est plus importante chez les MYB090 OE (environ 26 % chez les lignées 6 et 15, à plus de 40 % chez les lignées 19 et 23 par rapport au type sauvage) (figIII.16-A). Les plantes MYB090-SRDX OE ont montré une baisse de leur hauteur d'environ 22 % chez la lignée 1 et 30 % chez la lignée 5 (figIII.16-B). De même, la figure III.16-C illustre une baisse plus importante du diamètre des tiges des peupliers MYB090 OE. Le diamètre de la tige a baissé de 36 % chez les lignées 6 et 19 à environ 46 % chez les lignées 9 et 23 par rapport au sauvage. Chez les plantes MYB090-SRDX OE, la diminution du diamètre de tige a été moins importante comparée aux plants MYB090 OE. Cette diminution est d'environ 35 % chez la lignée 5. La lignée 1 quant à elle n'a pas montré une baisse significative de son diamètre de tige par rapport aux peupliers sauvages (figIII.16-D).

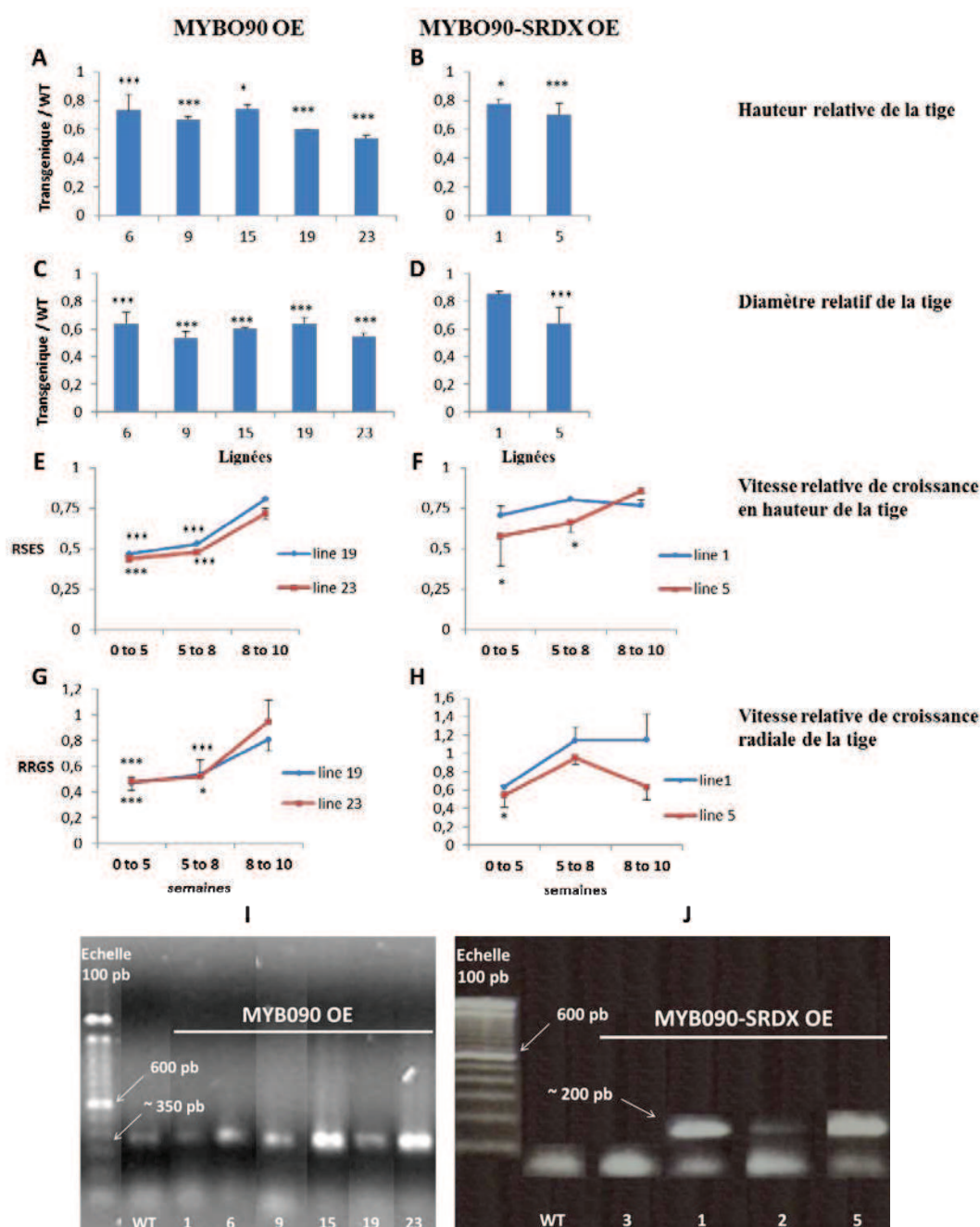


Figure III.16: Caractéristiques de croissance et expression du transgène chez les plantes MYB090 OE et MYB090-SRDX OE.

La hauteur (A, B) et le diamètre (C, D) de tige des peupliers transgéniques MYB090 OE ou MYB090-SRDX OE ont été rapportés, respectivement, à la hauteur et le diamètre de tige des plantes sauvages (WT). Les valeurs représentées sont ainsi le quotient des valeurs mesurées chez les plantes transformées divisées par les valeurs mesurées chez WT. La vitesse relative de la croissance en hauteur (RSES) (E, F) et la vitesse relative de croissance radiale (RRGS) (G, H) de la tige ont été calculées en divisant les valeurs mesurées chez les plantes transgéniques MYB090 OE (E, G) (lignée 19 : cercles bleus pleins ; lignée 23 : losanges rouges pleins) et MYB090-SRDXOE (F, H) (lignée 1 : cercles bleus pleins ; lignée 5 : losanges rouges pleins) sur les valeurs mesurées chez les plantes sauvages au niveau de 3 intervalles de croissance: jusqu'à 5 semaines (0 to 5), de la 5^{ème} semaine à la 8^{ème} semaine (5 to 8) et de la 8^{ème} semaine à la 10^{ème} semaine (8 to 10). Le niveau d'expression du transgène a été évalué par RT-PCR semi-quantitative (I, J). Les chiffres en dessous des bandes indiquent les numéros des lignées transgéniques. Les flèches indiquent le poids moléculaire. Les barres d'erreur représentent l'erreur type de la moyenne. Les différences significatives entre les valeurs correspondantes aux plantes transgéniques et celles du type sauvage sont représentés par (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

Ensuite, au contraire des plantes MYB221 OE, MYB221-SRDX OE et MYB156-SRDX OE pour lesquelles nous n'avons pas eu assez de données disponibles, nous avons pu analyser la vitesse d'élongation de la tige (RSES) et de la croissance radiale (RRGS) des peupliers MYB090 et MYB090-SRDX au niveau de 3 intervalles différents de la période de croissance.

La figure III.16-E et 16-G indiquent que, pendant 2 intervalles de croissance (de 0 à 5 semaines et de 5 à 8 semaines), les plants MYB090 OE ont montré une élongation de tige et une croissance radiale plus lentes en comparaison avec le type sauvage. En effet, durant les deux premiers intervalles de croissance (de 0 à 5 semaines et de 5 à 8 semaines), la vitesse de croissance de la tige chez les lignées 19 et 23 a été d'environ la moitié de celle mesurée chez les plantes sauvages et ceci que ce soit en hauteur ou en diamètre (figIII.16-E, G). Cependant, durant le troisième intervalle (de 8 à 10 semaines), la vitesse de croissance générale de la tige a augmenté d'une façon importante et a atteint un niveau similaire à celui de peupliers sauvages (figIII.16-E, G). Ces résultats montrent que les plants MYB090 OE ont, initialement, une croissance plus lente mais qui redevient normale après 8 semaines de croissance. Toutefois, l'influence de ce retard enregistré aux premières phases touche la taille et le diamètre de la tige qui restent tout de même plus petits par rapport au sauvage après 10 semaines de croissance.

Pendant les 5 premières semaines de croissance, les peupliers transgéniques surexprimant MYB090-SRDX ont montré une diminution d'environ 40 % de leur vitesse de croissance radiale en comparaison aux peupliers sauvages. Cependant, durant les deux intervalles suivants de croissance (entre 5 et 8 semaines et entre 8 et 10 semaines), la vitesse de croissance de diamètre de tige n'a pas été significativement différente de celle du témoin sauvage (figIII.16-H). L'élongation de tige des plantes MYB090-SRDX OE a été plus lente que chez les plantes sauvages durant les 2 premiers intervalles de croissance (entre 0 et 5 semaines et entre 5 et 8 semaines) et ceci chez la lignée 5 uniquement. La lignée 1 n'a montré aucune différence significative avec le type sauvage quel que soit l'intervalle de croissance (figIII.16-F). Ceci indique que la baisse de la vitesse de croissance de la tige chez les plantes MYB090-SRDX OE serait, durant la période testée, moins importante que chez les peupliers MYB090 OE.

Plusieurs autres modifications de caractéristiques phénotypiques ont été observées chez les plantes transformées que ce soit chez MYB090 OE ou MYB090-SRDX OE. Chez les peupliers MYB090 OE, la base de la tige est anormalement plus étroite que le reste de la tige (figIII.17-A). D'ailleurs, ces plantes ne sont pas capables de se tenir droites aisément sans tuteurage (figIII.17-B). Leurs entrenœuds sont plus courts que ceux du type sauvage (figIII.17-

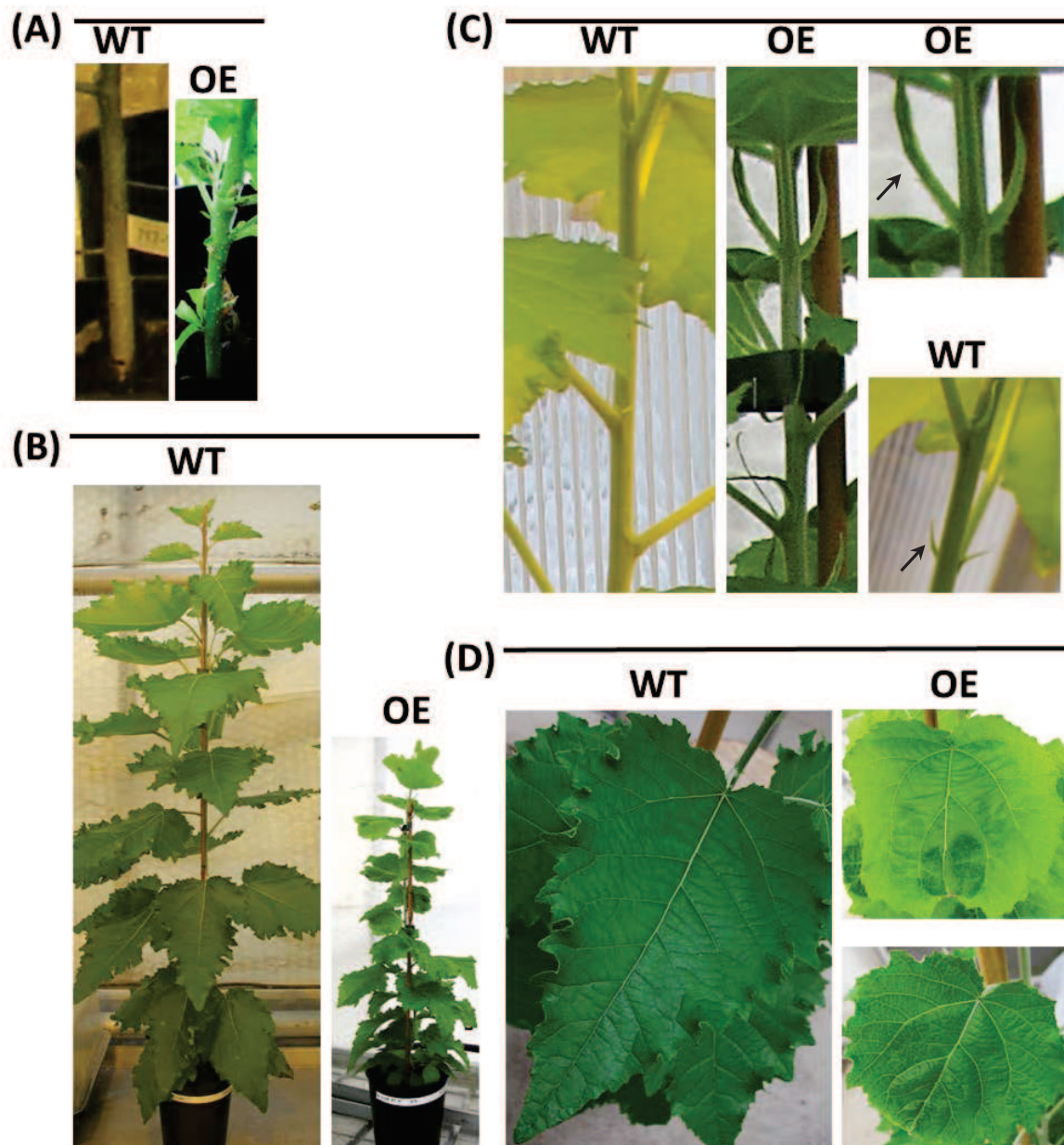


Figure III.17: Comparaison entre le phénotype des plantes MYB090 OE (OE) et les plantes sauvages (WT).

Le phénotype des plantes MYB090 OE (lignée 23) a été comparé avec celui des plantes sauvages (WT) au niveau de la plante entière (B), des feuilles (D), des entrenœuds (C), de la base de la tige (A) et des stipules (C). Les flèches montrent les stipules des transgéniques et du témoin (C).

C). Elles ont des feuilles plus petites, de couleur légèrement plus claire, déformées (moins allongées ou plus arrondies), et moins souples par rapport à celles du témoin (figIII.17-D). De plus, les stipules sont plus grands que ceux observés chez les peupliers sauvages (figIII.17-C). De même que les plantes MYB090 OE, les transgéniques MYB090-SRDX OE ont des entrenœuds plus courts (figIII.18-B) et ont une certaine difficulté à se tenir droit (figIII.18-A). Les feuilles de ces plantes sont enroulées vers le bas et légèrement plus petites que celles du type sauvage (figIII.18-C). Ces résultats suggèrent que la croissance et le développement des peupliers MYB090 OE sont peut-être plus touchés que ceux des plantes MYB090-SRDX OE.

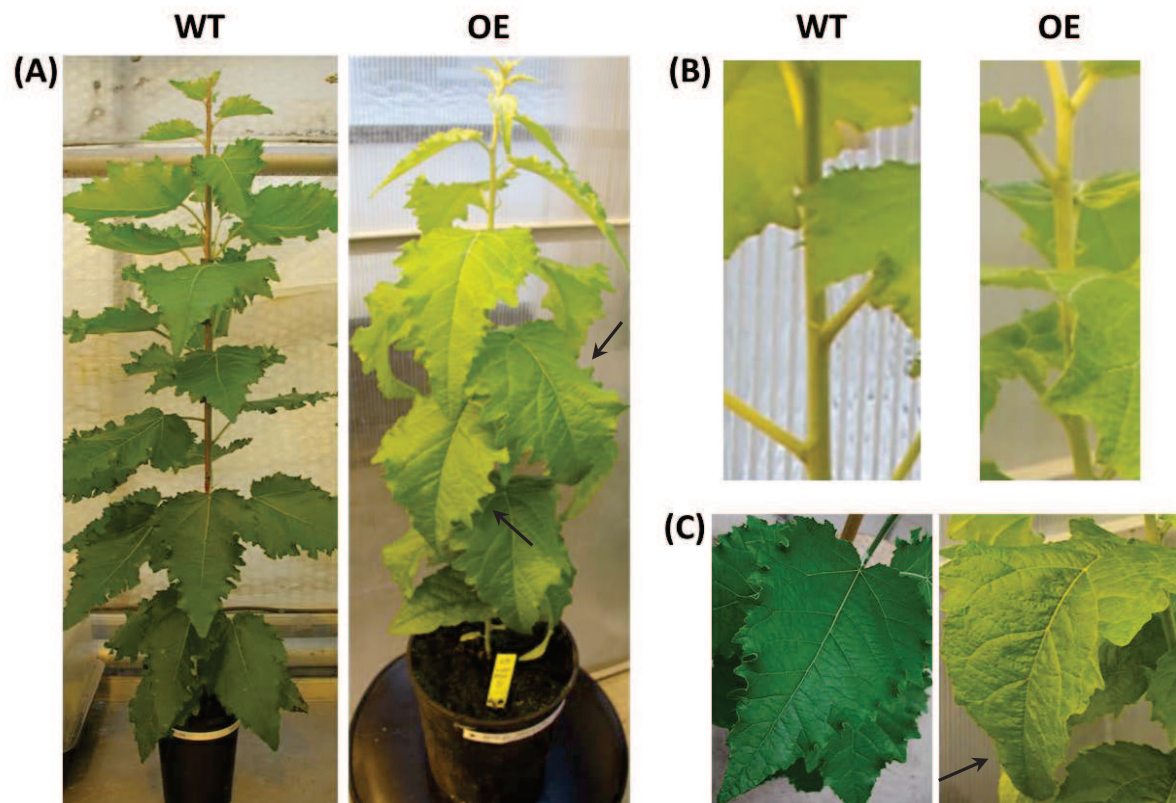


Figure III.18: Comparaison entre le phénotype des plantes MYB090-SRDX OE (OE) et les plantes sauvages (WT).

Le phénotype des plantes MYB090-SRDX OE (lignée 5) a été comparé avec celui des plantes sauvages (WT) au niveau de la plante entière (A), des feuilles (C) et des entrenœuds (B). Les flèches montrent les feuilles enroulées vers le bas des plantes transgéniques (A, C).

3.3.2- Caractéristiques histologiques

Les coupes transversales de tiges des plants MYB090-SRDX OE n'ont pas présenté de différences avec celles des plantes témoins (figIII.19-A,C; 20-C). Par contre, les observations de coupes colorées de tiges de plants MYB090 OE ont montré que si la majorité des fibres et des éléments de vaisseaux ont des parois lignifiées (coloration rouge) sauf les ponctuations (figIII.19-B ; 20-B), de nombreuses cellules associées aux vaisseaux (VAC) ont cependant des

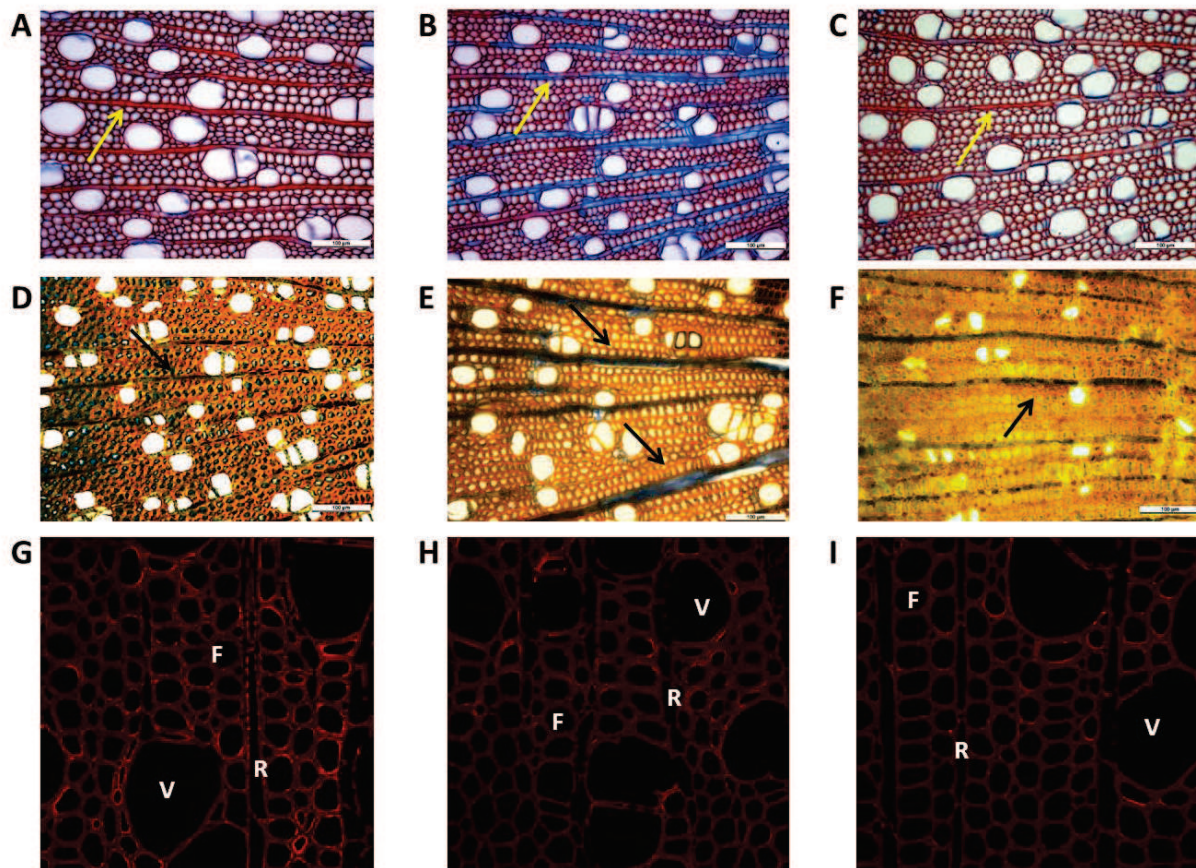


Figure III.19: Etude histologique des plantes MYB090 OE et MYB090-SRDX OE.

Les coupes transversales de tiges de peupliers sauvages non transformés génétiquement *Populus tremula x alba* clone INRA 717-1B4 (WT) (A, D, G), des plantes MYB090 OE (lignée 23) (B, E, H) et des plantes MYB090-SRDX OE (lignée 1) (C, F, I) ont été colorées par la Safranine-Bleu Alcian (A, B, C) ou par l'IKI-H₂SO₄ (D, E, F) et ont subi un immunomarquage des xylanes avec l'anticorps AX1 (G, H, I). Barre d'échelle = 100 µm. Agrandissements : (A-F) x20, (G-I) x40.

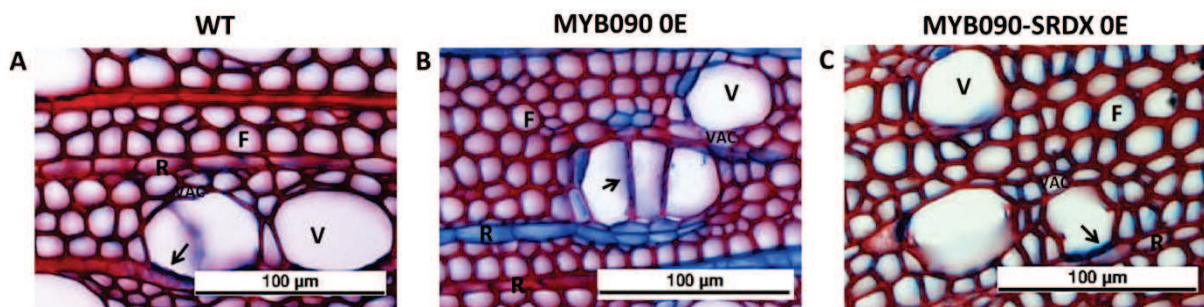


Figure III.20: Analyse histologique des plantes MYB090 OE et MYB090-SRDX OE.

Les coupes transversales de tiges de peupliers sauvages non transformés génétiquement *Populus tremula x alba* clone INRA 717-1B4 (WT) (A), des plantes MYB090 OE (B) et des plantes MYB090-SRDX OE (C) ont été colorées par la Safranine-Bleu Alcian. V, éléments de vaisseaux. F, fibres de xylème. R, rayons ligneux. VAC, cellules associées aux vaisseaux. Les flèches indiquent les ponctuations. Barre d'échelle = 100 µm.

parois moins lignifiées (figIII.20-B). De plus, plusieurs cellules de rayons de parenchyme, possèdent des parois colorées en bleu indiquant que ces dernières ne sont pas aussi lignifiées que les peupliers sauvages (figIII.20-B).

C'est pourquoi nous avons voulu savoir si les cellules de rayons des plants MYB090 OE contenaient plus de cellulose que chez les plantes sauvages, en utilisant une coloration avec de l'IKI/H₂SO₄ qui colore les tissus lignifiés en jaune/orangé et ceux riches en cellulose en bleu foncé. Il est à noter que les parois cellulaires se gonflent au fur et à mesure sous l'effet de l'acide sulfurique. Dans ce cas, il est possible de voir la cellulose au niveau des tissus lignifiés ce qui résulte en une coloration verdâtre. Après quelques minutes de l'application de cette coloration sur les coupes de tige, les cellules finissent par éclater. De ce fait, cette coloration n'est pas stable dans le temps, ce qui explique les éventuelles différences entre les coupes au niveau de l'intensité de la coloration et l'aspect des cellules. Les coupes de tige des plantes témoins, colorées à l'aide de cette technique, ont montré une coloration orangée des parois des fibres, des éléments de vaisseaux et des rayons indiquant une lignification normale de ces parois (figIII.19-D). La partie interne de plusieurs cellules de parenchyme est colorée en noir (figIII.19-D). Ceci pourrait indiquer la présence d'amidon dans les cellules de rayons car l'IKI (utilisé dans cette coloration) colore l'amidon en noir. Il est à noter aussi que quelques cellules de rayons ont été colorées en vert suggérant une présence simultanée de cellulose et de lignines au niveau de leurs parois (figIII.19-D). Chez les coupes de tige des plants MYB090 OE, les parois des fibres et des éléments de vaisseaux ont une coloration similaire à celle observée chez les plantes sauvages. Cependant, plusieurs cellules de rayons de parenchyme, au niveau de certaines zones de la tige, ont une couleur bleu foncée (figIII.19-E) indiquant que leur paroi est riche en cellulose. Certaines cellules de rayons ont été colorées en noir (figIII.19-E), probablement, à cause de la présence de l'amidon. Nous avons également coloré de coupes transversales de tiges de plantes MYB090-SRDX OE à l'aide de la même technique sans avoir décelé de différences avec les coupes des peupliers sauvages (figIII.19-F).

MYB090 est un orthologue des FT MYB54 et MYB52 connus comme étant des activateurs des gènes de biosynthèse des xylanes chez *Arabidopsis thaliana*. C'est pourquoi nous avons posé l'hypothèse que MYB090 pouvait être un régulateur positif des gènes de biosynthèse de xylanes chez le peuplier. Nous avons donc réalisé une expérience d'immunolocalisation des arabinoxylanes sur des coupes transversales des plantes transgéniques en utilisant l'anticorps AX1 (Guillon *et al.* 2004). Les images obtenues montrent que, chez les plantes MYB090 OE et MYB090-SRDX OE, la fluorescence due à

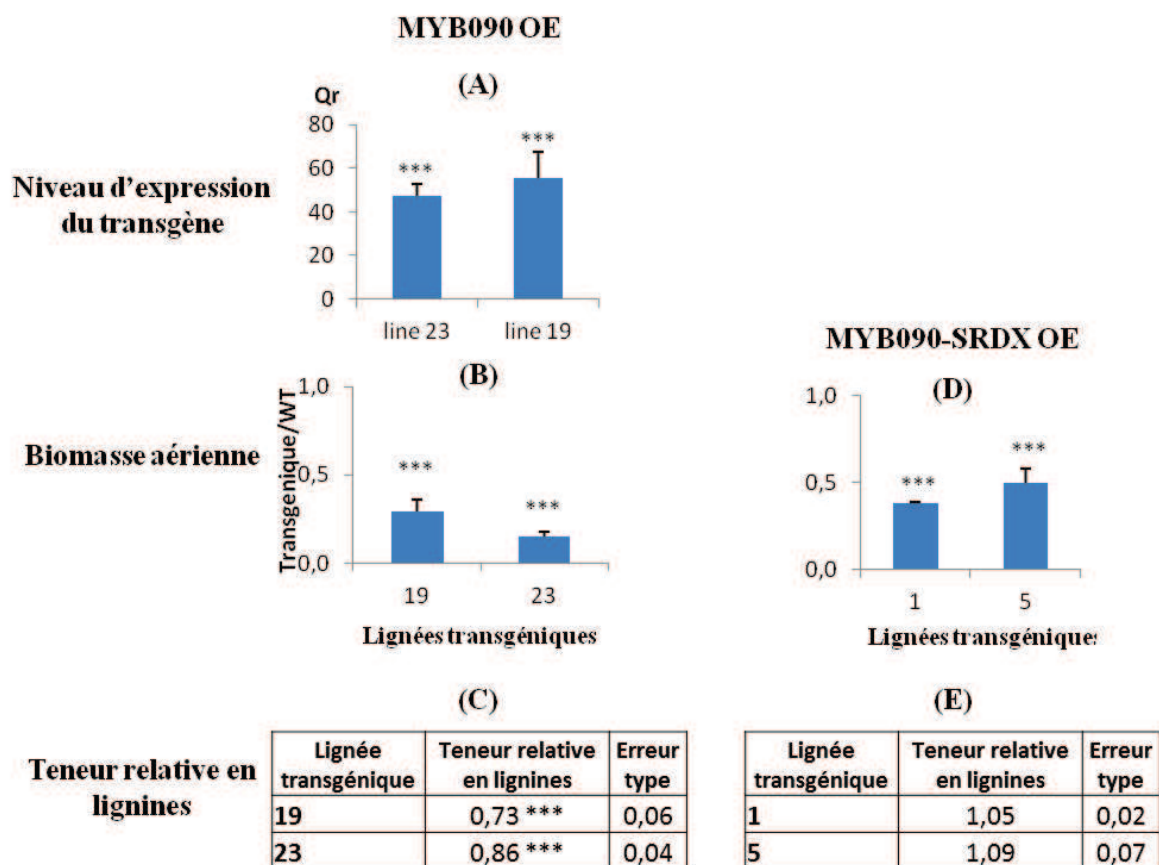


Figure III.21: Mesure du taux de lignines totales, de biomasse et d'expression du transgène chez les peupliers MYB090 OE et/ou MYB090-SRDX OE.

Le taux de lignines totales a été mesuré par la méthode Klason (C, E). La partie aérienne des plantes (tige + feuilles) a été pesée pour évaluer la biomasse aérienne (B, D). Le taux de lignines totales et le poids de la tige ont été rapportés aux valeurs mesurées chez les plantes sauvages. Qr représente l'expression quantitative de *MYB090* (A) rapportée à celle mesurée chez le sauvage. Les barres d'erreur représentent l'erreur type de la moyenne. Les différences significatives entre chaque lignée transgénique et le témoin sont représentées par : (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

Tableau III.2 : Taux relatif d'holocellulose et d' α -cellulose dans les plantes MYB090-SRDX OE par rapport aux plantes sauvages

Transformation	Lignée	Taux relatif d'holocellulose	Erreur type	Taux relatif d' α -cellulose	Erreur type
MYB090-SRDX	1	0,99	0,021	0,95	0,009
MYB090-SRDX	5	1,00	0,023	0,96	0,017

l'immunomarquage des arabinoxyanes est légèrement plus faible que celle observée chez le témoin. De plus, aucune différence n'a été remarquée entre les deux types de peupliers transgéniques (figIII.19-G, H, I). Ceci laisse penser que MYB090 ne semble pas être un activateur des gènes de biosynthèse des arabinoxyanes, reconnues par AX1, chez les peupliers.

3.3.3- Analyses biochimiques et de biomasse

Nous avons mesuré la teneur des plantes MYB090 OE et MYB090-SRDX OE en lignines totales. Les résultats montrent que chez les plantes transgéniques, une baisse significative du taux de lignines totales a été enregistrée. Cette baisse est d'environ 14 % et 26 %, respectivement, chez les lignées 23 et 19 par rapport aux peupliers sauvages (figIII.21-C). Elle est accompagnée d'une forte expression du transgène (aux alentours de 50 fois plus que chez le sauvage) chez ces deux lignées (figIII.21-A). Les peupliers transformés ont montré également une très forte baisse de leur biomasse aérienne allant de 70 % à 85% respectivement chez les lignées 19 et 23 (figIII.21-B). Quant aux plantes MYB090-SRDX OE, elles n'ont pas montré de différences avec les peupliers sauvages en ce qui concerne le taux de lignines totales (figIII.21-D). Toutefois, elles ont enregistré une baisse importante de leur biomasse aérienne qui allait de 50 % à 60 % respectivement chez les lignées 5 et 1 (figIII.21-E).

Nous avons voulu vérifier aussi si les plantes transgéniques sur-exprimant *MYB090* ont plus de cellulose que les plantes sauvages. Donc nous avons procédé à une analyse biochimique du taux de cellulose chez les plantes transgéniques. Malheureusement, cette analyse n'est pas adaptée à de faibles quantités de poudre de tiges broyées. Nous n'avons pu mesurer avec certitude le taux de cellulose que sur les plantes sauvages et celles sur-exprimant *MYB090-SRDX*. Ces derniers n'ont pas montré de différences avec le témoin que ce soit pour leur teneur en holocellulose ou en alpha-cellulose (tabIII.2).

3.3.4- Expression de gènes de biosynthèse des lignines, cellulose et xylanes

La quantification du niveau de transcrits a montré que deux gènes parmi ceux intervenant dans la biosynthèse de cellulose (*CESA4* et *CESA8*) sont régulés positivement chez les plantes MYB090 OE (figIII.22). L'expression d'un des gènes de biosynthèse des xylanes *IRX8* est très légèrement réprimée chez les plantes transgéniques tandis que celle du deuxième gène (*IRX9*) ne change pas par rapport aux plantes sauvages (figIII.22). En ce qui concerne les

gènes de biosynthèse de lignines, le niveau d'expression de deux d'entre eux (*PAL1* et *CCR1*) a presque doublé dans le xylème des plantes transformées en comparaison avec le type sauvage. Cependant celle du gène *CAD* est régulée négativement (figIII.22).

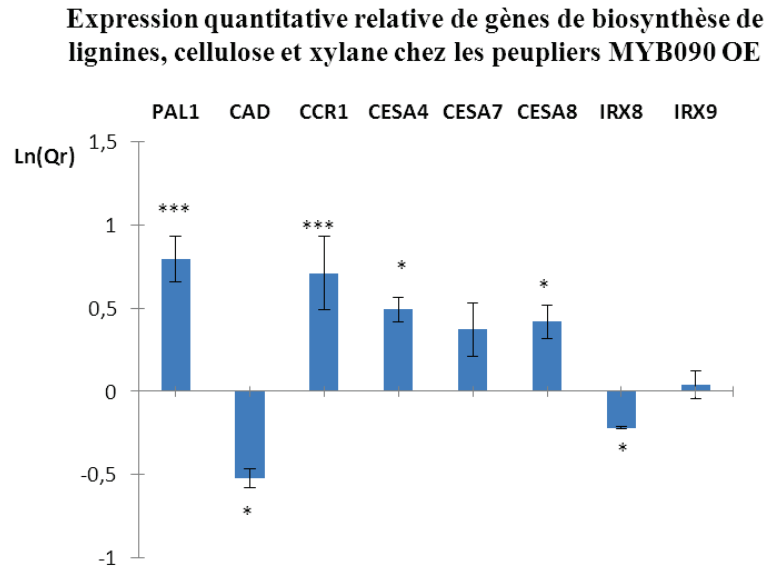


Figure III.22: Régulation de gènes de biosynthèse des lignines, cellulose et xylanes dans le xylème des plantes MYB090 OE.

Le logarithme des valeurs de l'expression quantitative relative ($\text{Ln}(\text{Qr})$) des gènes de biosynthèse de lignines (*PAL1*, *CAD*, *CCR1*), cellulose (*CESA4*, *CESA7*, *CESA8*) et xylanes (*IRX8*, *IRX9*) au niveau du xylème de plantes transgéniques surexprimant *MYB090* est présenté par des rectangles bleus. Les barres d'erreur représentent l'erreur type de la moyenne. La différence significative entre les valeurs d'expression des plantes transgéniques et celles des plantes sauvages est représentée par (* t-test $P < 0.05$) ou (***) t-test $P < 0.01$).

Expression de 8 gènes cibles de MYB090 dans le xylème des plantes MYB090 OE ou MYB090-SRDX OE

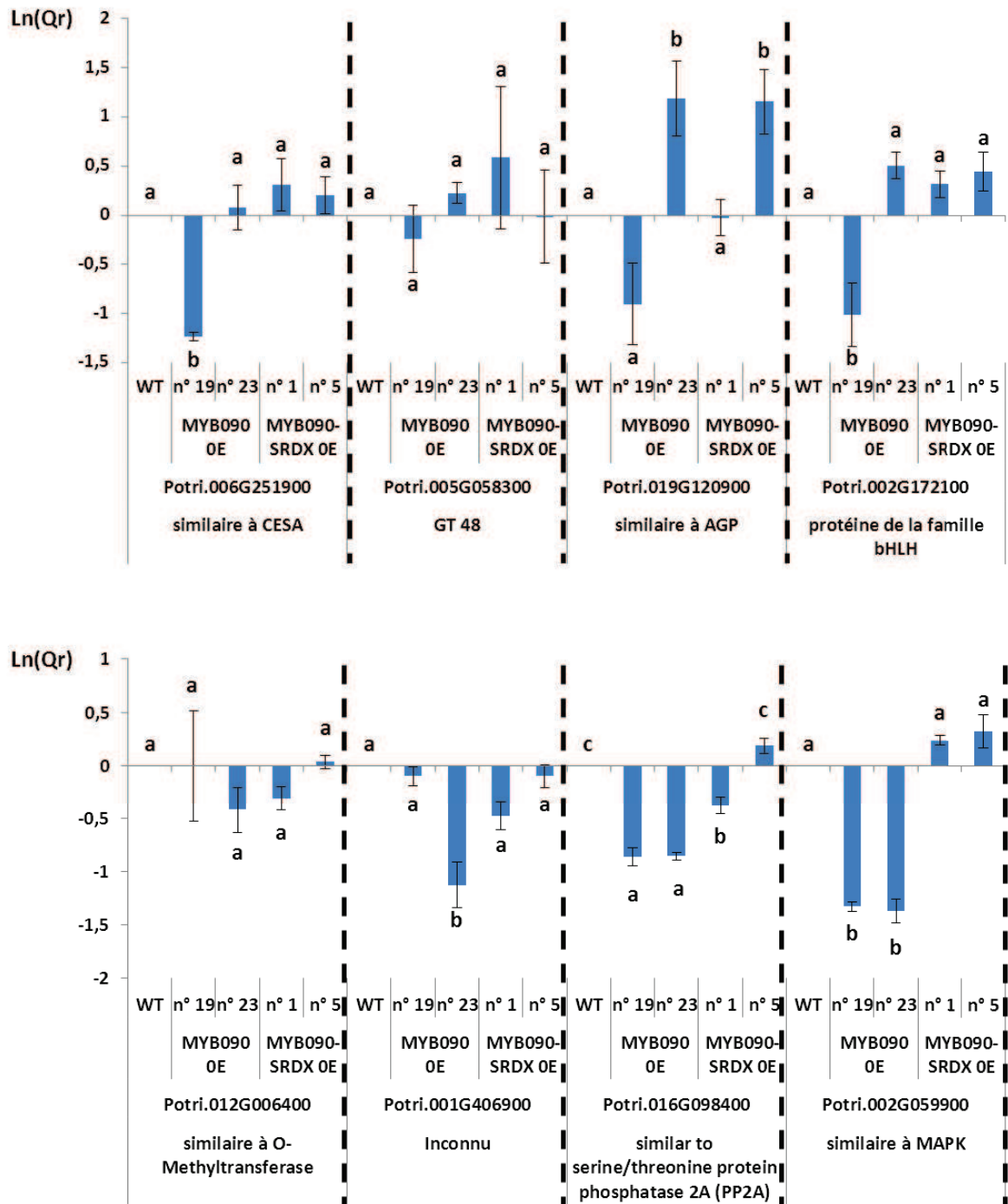


Figure III.23: Expression de 8 gènes cibles de MYB090 dans le xylème des plantes MYB090 OE et MYB090-SRDX OE.

Le logarithme des valeurs de l'expression quantitative relative ($\ln(Qr)$) de 8 gènes cible de MYB221 identifiés par ChIP-Seq (voir modèle de gène et annotation au-dessous du graphique) au niveau du xylème de plantes transgéniques surexprimant *MYB090* (MYB090 OE) ou *MYB090-SRDX* (MYB090-SRDX OE). Les barres d'erreur représentent l'erreur-type de la moyenne. La différence significative du niveau d'expression ($\ln(Qr)$) entre les différentes lignées y compris les plantes sauvages (WT) est représentée par des lettres différentes (t-test $P < 0.01$) pour chaque gène cible séparément.

3.3.5- Expression de gènes cibles identifiés par ChIP-SEQ

L'expression de 8 gènes sélectionnés parmi les cibles potentielles de MYB090 identifiées par ChIP-SEQ a été étudiée chez les plantes transgéniques surexprimant MYB090 ou MYB090-SRDX. Les études de qRT-PCR montrent qu'il existe une certaine variabilité entre les lignées transgéniques au sein d'une même transformation. Toutefois, l'expression de deux cibles potentielles (gène similaire à MAPK et gène similaire à PP2A) est réprimée chez les plantes MYB090 OE et quasiment inchangée par rapport au témoin non transgénique chez les peupliers MYB090-SRDX (figIII.23). De même, l'expression d'autres gènes cibles (gène similaire à CESA, gène codant pour une protéine de la famille bHLH, et gène codant pour une protéine de fonction inconnue) est réprimée chez les plantes MYB090 OE sans qu'elle soit modifiée chez les plantes MYB090-SRDX OE. Toutefois, cette répression n'est observée uniquement chez une seule lignée pour chaque cas (lignée 19 pour le gène similaire à CESA, protéine de la famille bHLH et lignée 23 pour le gène de fonction inconnue). Pour la cible GT48 et celle similaire à une O-méthyltransférase, aucune différence d'expression n'a été observée que ce soit chez les plantes transgéniques ou les témoins (figIII.23). Le niveau d'expression de la cible similaire à AGP semble augmenter chez les deux types de plantes transgéniques mais uniquement chez une seule lignée pour chaque type (figIII.23).

4- Discussion

Dans cette étude, nous avons choisi trois facteurs de transcription de peuplier de la famille des MYB. MYB221 et MYB156 appartiennent au sous-groupe 4 de la famille MYB et sont potentiellement des répresseurs de la voie de biosynthèse de lignines ; MYB090 est proche d'*AtMYB52* et *AtMYB54*, dont l'action activation/répression est moins claire. De façon systématique, nous avons suivi une approche de transgénèse consistant en deux stratégies différentes : (1) la surexpression de *MYB090*, *MYB221* et *MYB156* natifs afin d'étudier *in planta* l'effet de cette surexpression, (2) la surexpression de la forme répressive dominante des 3 FT fusionnés au motif SRDX. Cette stratégie est surtout intéressante dans le cas de FT activateurs de transcription. Cependant, une telle stratégie a été utilisée sur KNAT7 qui est un répresseur de transcription chez le peuplier : l'effet répresseur de KNAT7 est accru après l'intégration du motif SRDX (Li *et al.* 2012). Nous verrons que c'est effectivement le cas pour nos propres essais.

La surexpression de *MYB156* et de *MYB221-SRDX* a engendré une baisse du taux de lignines par rapport aux peupliers sauvages. La transformation a donné lieu aussi à une diminution très importante de la lignification de toutes les parois des cellules du xylème (fibres de xylème et rayons de parenchyme) à l'exception des parois des éléments de vaisseaux. Ceci s'est accompagné d'une baisse considérable de la hauteur et du diamètre des plantes transgéniques avec une formation de feuilles plus petites et légèrement déformées (en forme de cuvette). Tous ces résultats sont similaires à ceux obtenus suite à la surexpression d'*EgMYB1* chez le peuplier et chez *Arabidopsis* (Legay *et al.* 2010) et de *ZmMYB31* chez *Arabidopsis* (Fornale *et al.* 2010). La surexpression de *CmMYB1*, un FT de chrysanthème appartenant au sous-groupe 4 de la famille MYB et orthologue d'*EgMYB1*, chez *Arabidopsis* a engendré une baisse de la lignification des parois ainsi qu'une baisse significative du taux de lignines chez les plantes transformées (Zhu *et al.* 2013). De plus, plusieurs gènes de biosynthèse de lignines (*C3H*, *4CL*, *F5H*, *COMT*, *PAL*, *CCoAOMT*, *CCR*, *CAD*) (Legay *et al.* 2010 ; Fornale *et al.* 2010 ; Zhu *et al.* 2013) sont régulés négativement dans les plantes transgéniques. En accord avec ces observations, l'expression de deux gènes de la voie de biosynthèse des lignines (*CAD* et *CCR1*) est réprimée chez nos plantes transformées. D'ailleurs, deux gènes d'*Eucalyptus* (*EgCAD2* et *EgCCR*) impliqués dans la voie de biosynthèse des lignines sont régulés directement et négativement par *EgMYB1* (Legay *et al.* 2007). Allant dans le même sens, le 'silencing' de *CCR* chez le peuplier et le tabac a donné lieu à des plantes plus petites avec des feuilles déformées, ayant des parois cellulaires moins lignifiées et un taux en lignines plus faible (Leplé *et al.* 2007; Piquemal *et al.* 1998). *PAL1* est toutefois légèrement régulé positivement chez les plantes *MYB156* OE. *PAL1* catalyse la première étape de la biosynthèse des monolignols *via* la voie des phenylpropanoïdes. L'effet de sa légère activation chez les plantes *MYB156* OE a été probablement contré par l'effet de la répression de *CAD* et de *CCR1* intervenant au niveau des étapes finales de la formation des monolignols. De plus, certains travaux réalisés sur des plantes sur-exprimant des FTs MYB répresseurs de la lignification montrent que les gènes de biosynthèse de lignines ne sont pas tous régulés négativement (Ma *et al.* 2011 ; Fornale *et al.* 2010 ; Legay *et al.* 2010 ; Jin *et al.* 2000). Il est d'ailleurs intéressant de noter que les parois des vaisseaux, riches en lignines de type G (Saito *et al.* 2012) semblent moins affectées, ce qui laisse à penser que la voie de biosynthèse des unités S, passant notamment par la *F5H* et la *COMT*, est plus perturbée par la modification. D'autres gènes impliqués dans la biosynthèse de deux autres polymères principaux de la paroi cellulaire (cellulose et hémicelluloses) sont régulés négativement chez les peupliers sur-exprimant

MYB156, ce qui est similaire aux résultats de Legay *et al.* (2010) et Zhu *et al.* (2013) indiquant un rôle de *MYB156*, non seulement dans la biosynthèse des lignines mais aussi dans la régulation de la formation de la paroi secondaire d'une façon générale. Ainsi, *MYB156* et *MYB221-SRDX* semblent avoir les mêmes fonctions au niveau de la régulation de la lignification des parois secondaires et de la croissance des plantes.

Il est à noter que la lignée 15 des plantes *MYB156* OE a présenté une modification très importante de la lignification des parois secondaires des fibres et des rayons ainsi qu'une baisse considérable du taux de lignines totales. Elle a cependant présenté une vitesse de croissance relative en constante augmentation, et supérieure à celle du sauvage. Ceci suggère que pour des échantillonnages plus tardifs (par exemple 6 mois, au lieu de 3 mois comme dans ce travail), cela aurait pu changer les données de croissance et de biomasse. Ce sont certainement des données intéressantes à acquérir pour identifier éventuellement des lignées intéressantes en terme de teneur en lignines et production de biomasse pour une exploitation à des fins de biocarburants.

MYB221 doit probablement avoir la même fonction que *MYB156* mais on n'observe pas de phénotype similaire chez les plantes *MYB221* OE. Ceci pourrait être dû au niveau d'expression du transgène qui est relativement faible chez ces plantes transgéniques et qui est proche de celui des contrôles transgéniques des autres transformations. En revanche, si on interprète les résultats de surexpression de *MYB221* et de *MYB221-SRDX* seuls, on pourrait considérer que *MYB221* serait plutôt un activateur de la lignification puisque la répression dominante de ce FT a fortement inhibé la lignification et la croissance des plantes transformées. L'absence d'un phénotype de 'sur-lignification' des parois ou de 'sur-croissance' des plantes peut être expliquée aussi par le faible niveau d'expression du transgène. Cependant, *MYB221* possède un motif répresseur, mais en même temps semble exercer une fonction d'activation car l'une des cibles potentielles directes de *MYB221* identifiée par ChIP-SEQ (alcool déshydrogénase) est régulée positivement chez les plantes *MYB221* OE. Les autres cibles sont par contre régulées négativement. Il est intéressant de noter qu'une des cibles potentielles de *MYB221* qui est un FT MYB, appartenant au même groupe de MYB répresseur de la lignification (Jin *et al.* 2000) est régulée négativement chez ces plantes. Dans sa séquence, on retrouve le motif répresseur LxLxLx de répression. Ceci laisse penser à une éventuelle boucle de rétrorégulation ou rétrocontrôle si on pose l'hypothèse que *MYB221* réprime un répresseur de la lignification. Cela expliquerait pourquoi on n'observe pas de phénotype de baisse de

lignification chez les plantes MYB221 OE. Cette boucle de rétro-contrôle ne fonctionnerait pas avec MYB221-SRDX.

Nous avons observé une gradation dans le phénotype entre les plantes surexprimant MYB221 et MYB156 d'une part et leur formes répressives dominantes MYB221-SRDX et MYB156-SRDX. Si on suppose que ces deux FT sont des répresseurs, la répression dominante aurait amplifié le phénomène de répression (on a par exemple un motif de répression en plus et une plus grande quantité de transcrits car le FT-SRDX est sous contrôle d'un promoteur 35S fort), ce qui donnerait un phénotype plus prononcé. Ce phénomène est observé chez Li *et al.* (2012) dans le cas du FT KNAT7.

La surexpression de *MYB156* fusionné au domaine de répression dominante SRDX a engendré des peupliers nains présentant une mortalité importante en serre et ayant une croissance très lente ainsi que des feuilles très petites et déformées. Ces plantes ont développé des racines gigantesques (phénomène observé *in vitro* sans pouvoir le vérifier en serre). D'ailleurs, les données d'expression de l'eFP Browser (Wilkins *et al.* 2008) montrent une expression assez prononcée de MYB156 au niveau de racines. Ceci suggère peut-être un rôle important de ce FT dans la croissance racinaire ou peut-être dans la répartition des ressources au sein des organes de la plante affectant l'équilibre entre les parties aérienne et racinaire.

Les peupliers surexprimant *MYB221* ont montré un développement d'un nombre plus important de branches sylleptiques chez les plantes transgéniques. Ce phénotype a été observé chez des peupliers surexprimant un gène de châtaignier *CsRAVI* (Moreno-Cortés *et al.* 2011). Ceci nous laisse penser à une relation possible entre MYB221 et des éventuels orthologues de *CsRAVI* chez le peuplier, que ce soit à travers une régulation directe ou indirecte.

Nous avons également voulu caractériser fonctionnellement le FT MYB090, proche de MYB52 et MYB54 chez *Arabidopsis*, connus pour leur rôle dans la régulation positive de certains gènes de biosynthèse de cellulose, xylanes et lignines ainsi que dans le développement de la paroi cellulaire (Zhong *et al.* 2008). Nous avons également choisi ce FT car il est surexprimé dans le bois de tension riche en cellulose et pauvre en lignines. Les plantes surexprimant *MYB090* ont toutefois présenté une diminution de la lignification d'un nombre de cellules associées aux vaisseaux mais aussi de la majorité des cellules de rayons de parenchyme, alors que la lignification des fibres ne présente pas de différence avec le témoin. Les peupliers transgéniques ont montré une baisse significative de leur teneur en lignines totales, alors que seuls les rayons étaient touchés. Nous avons réalisé la totalité des études histologiques au niveau

de la base de la tige, c'est-à-dire là où les cellules sont les plus âgées, alors que les dosages de lignines ont été faits sur un fragment de tige allant d'un peu au-dessus de la base jusque la première feuille pleinement étalée. Ce fragment de tige est moins lignifié que chez le sauvage. On peut faire l'hypothèse qu'il y a un retard de lignification des cellules jeunes du xylème au niveau de la tige, qu'on ne peut pas observer à la base de celle-ci.

L'expression de la majorité des cibles potentielles de MYB090 testée par qPCR est réprimée d'une façon générale, dans les MYB090 OE. Ceci est assez concordant avec les travaux de Karpinska *et al.* 2004 qui proposent que PttMYB021 (homologue de MYB090 chez *Populus tremula x tremuloides*) serait un répresseur transcriptionnel et que le gène *CCoAOMT* est une cible potentielle de ce FT. D'ailleurs, nous avons pu identifier par ChIP-SEQ au moins un gène *CCoAOMT* parmi les cibles potentielles de MYB090. Ainsi, nous postulons que cette répression pourrait être plus forte au niveau des cellules de rayon. Toutefois, chez les plantes surexprimant la forme de répression dominante, aucune différence au niveau histologique ni biochimique (teneur en lignines) n'est constatée avec le témoin alors que la répression dominante de *MYB52* et *MYB54* chez *Arabidopsis* a abouti à une diminution de l'épaisseur de la paroi de cellules de fibres et de vaisseaux (Zhong *et al.* 2008).

Certaines cellules de parenchyme possèdent même des parois cellulaires probablement plus riches en cellulose. D'ailleurs, les gènes *CESA4* et *CesA8* sont légèrement activés chez les plantes MYB090 OE. De même, un gène de la voie de biosynthèse de cellulose est régulé positivement suite à la surexpression de *MYB52* et *MYB54* chez des protoplastes de feuilles d'*Arabidopsis* (Zhong *et al.* 2008). De plus, nous avons identifié plusieurs gènes de biosynthèse de cellulose parmi les cibles potentielles de MYB090 (Voir chapitre 2). Cependant, il pourrait s'agir uniquement d'une perte de la lignification des parois secondaires des rayons rendant la cellulose plus visible après sa coloration.

Les peupliers surexprimant *MYB090* ont montré également un raccourcissement des entrenœuds, une déformation des feuilles et de la forme de la tige et un développement très important de stipules. Un phénotype semblable a été observé chez des peupliers surexprimant un micro ARN de maïs de la famille *MIR156* appelé *Corngrass1(Cg1)* sous le contrôle de 35S. Ces plantes ont montré de la même façon une baisse de leur teneur en lignines (Rubinelli *et al.* 2012). Cette forte ressemblance entre les deux cas peut être expliquée par une éventuelle interaction entre MYB090 et *Cg1*, l'un étant sous le contrôle de l'autre ou inversement. Toutefois, certains effets liés au fait que la surexpression est constitutive dans les deux cas

n'écarter pas l'hypothèse de modifications non liées aux fonctions de *MYB090* ou *Cgl* mais plutôt résultants de la perturbation causée par la surexpression de ces gènes dans des organes où leur expression pourrait être absente en conditions naturelles.

La surexpression de *MYB090* a résulté en une légère baisse de l'immunomarquage des arabinoxylanes chez les plantes *MYB090* OE et *MYB090-SRDX* OE en comparaison aux plantes sauvages. Conformément à ces observations, l'expression du gène *IRX8* est légèrement inhibée chez les peupliers *MYB090* OE alors que celle de *IRX9* est inchangée. Ainsi, *MYB090*, bien qu'orthologue de *MYB52* et *MYB54* ne semble pas avoir les mêmes rôles que ces derniers (activateurs de la synthèse des xylanes). Il a un rôle plus complexe touchant d'une part le développement des parois cellulaires de certains types de cellules de xylème, d'autre part la croissance de la plante de façon générale.

Discussion générale

Pour caractériser fonctionnellement 3 FTs de peuplier (MYB090, MYB221 et MYB156) potentiellement impliqués dans la formation du bois d'une façon générale, et plus précisément de la paroi secondaire, nous avons cherché à identifier leurs cibles, car elles constituent la clé pour comprendre leur fonction. Pour cela, nous avons utilisé une approche novatrice chez les plantes, le ChIP-SEQ, qui permet de déterminer *in vivo* les cibles d'un FT à l'échelle du génome. Cette approche a été mise en œuvre pour MYB221 et MYB090. En parallèle, nous avons surexprimé chacun de ces deux FT ainsi que leurs formes répressives dominantes (SRDX) chez des plants de peuplier afin d'étudier *in planta* l'effet de ces transformations sur le phénotype. Dans notre étude, MYB156 a été étudié uniquement par transgénèse.

L'étude de transgénèse réalisée dans ma thèse présente l'avantage d'avoir transformé des plants de peuplier avec des FT clonés chez la même espèce (*Populus tremula* x *alba*). Ainsi, l'interprétation des phénotypes observés devrait être facilitée par le fait que les interactions géniques et protéiques existent déjà dans les cellules, plutôt que s'il s'agissait de l'expression d'un gène hétérologue. Toutefois, nous avons surexprimé les FT étudiés sous le contrôle d'un promoteur constitutif fort. Donc, le transgène sera exprimé de façon élevée dans tous les organes de la plante, y compris les organes où le FT endogène n'est pas ou peu exprimé naturellement : le FT, alors abondant dans les cellules transgéniques, peut interagir avec des acteurs biologiques qui n'ont pas naturellement de relation avec lui au niveau fonctionnel, engendrant ainsi des éventuels déséquilibres, et donc des phénotypes qui n'ont pas obligatoirement de lien avec la fonction du FT, d'où un risque de biais lors de l'interprétation des résultats.

MYB156 est un répresseur direct ou indirect de la lignification

Les plantes surexprimant *MYB156* présentent une baisse de la teneur en lignines totales de leurs tiges, associée à une diminution importante de la lignification des parois secondaires des fibres et des rayons (coloration à la Safranine/Bleu Alcian et auto-fluorescence, figIII.13-C, D ; figIII.14-B). Par ailleurs, l'expression de plusieurs gènes de biosynthèse des lignines mais aussi de la cellulose et des xylanes est réprimée chez ces plantes (figIII.15). Cette diminution de la lignification s'accompagne d'une diminution de croissance, en hauteur et en diamètre des tiges (figIII.9-B, E) ainsi que d'une modification de la forme des feuilles, ce qui est généralement observé lorsque l'on modifie le métabolisme des lignines

(Legay *et al.* 2010 ; Fornale *et al.* 2010). MYB156 possède le motif de répression de type EAR. Au vu de ces différents éléments, nous pouvons conclure que MYB156 semble bien avoir des fonctions très similaires à celles d'EgMYB1. C'est un répresseur de la lignification dans les fibres et les rayons du xylème.

Les plantes surexprimant *MYB156-SRDX* présentent un phénotype beaucoup plus affecté puisque les plantes sont naines, dès l'étape de culture *in vitro*, et ont du mal à passer le cap de l'acclimatation en serre. Les teneurs en lignines n'ont pas été déterminées sur ce matériel, mais il est vraisemblable qu'elles soient fortement diminuées, impactant la croissance de la plante de façon notable. MYB156-SRDX possède 2 motifs de répression de la transcription, un motif natif et un motif ajouté par construction en position C-terminale. Comme observé pour KNAT7 (Li *et al.* 2012 ; Zhong *et al.* 2008), on observe alors un effet répresseur plus fort puisque des plantes surexprimant KNAT7 fusionné au motif de répression EAR ont présenté des parois secondaires plus fines que des plantes surexprimant KNAT7.

MYB156 est donc un répresseur de la lignification des parois, mais on ne sait pas, à ce stade, si c'est un répresseur direct ou indirect des gènes de biosynthèse de lignines. Afin de répondre à cette question, on peut chercher à identifier les cibles génomiques directes de MYB156 par ChIP-SEQ sachant que dans ce cas, il sera nécessaire de produire un anticorps spécifique pour MYB156. Alternativement, on peut aussi envisager des expériences *in vitro* telles que le simple hybride pour tester l'interaction entre MYB156 et un petit nombre de promoteurs de gènes candidats. On pourra aussi vérifier si MYB156 est un répresseur de la lignification uniquement ou s'il est impliqué dans la régulation d'autres métabolismes impliqués dans la formation de la paroi secondaire.

Pour la production de biocarburants à partir de la matière ligno-cellulosique du bois, il est intéressant d'avoir du bois moins riche en lignines, polymère que l'on doit éliminer pour avoir accès à la cellulose. Cependant, il faut que les plants puissent avoir une production de biomasse suffisante. Il est à noter que la lignée n°15 des plants surexprimant MYB156 montre à 12 semaines d'élevage en serre une baisse d'environ 25 % au niveau de la teneur en lignines (figIII.6-G) associée à une biomasse plus faible (figIII.6-D). Cette lignée a cependant une vitesse de croissance de plus en plus rapide, que ce soit en hauteur ou en diamètre (figIII.10), montrant que si la phase d'acclimatation et de démarrage de croissance a été difficile, la lignée tend à rattraper son retard de croissance. Une durée de culture plus longue (par exemple 6 mois ou plus) suivie de mesures de biomasse et du taux de lignines permettrait de voir si cette lignée peut produire plus de biomasse avec moins de lignines. Si c'est le cas, elle

serait intéressante pour l'industrie des biocarburants. Par ailleurs, nous pouvons envisager des tests de microgreffage *in vitro* des tiges de cette lignée sur les racines très développées des plantes MYB156-SRDX OE dans un but d'améliorer encore plus leur croissance. En effet, les plantes greffées auront des racines plus développées qui vont favoriser peut-être la croissance de la tige MYB156 OE qui est moins lignifiée. Ce qui pourrait donner des peupliers avec une biomasse plus importante mais aussi contenant moins de lignines, d'où un intérêt éventuel pour l'industrie des biocarburants.

MYB221 est un répresseur de la lignification contrôlant probablement ses gènes cibles *via* un motif SMRE

MYB221 est un orthologue d'EgMYB1, connu pour être un FT répresseur de la lignification (Legay *et al.* 2010). Sa séquence protéique contient le motif EAR de répression de la transcription. Nous avons donc posé l'hypothèse que MYB221 a une fonction similaire à EgMYB1, dont 2 cibles (CAD et CCR) ont été démontrées expérimentalement (Legay *et al.* 2007). Nous avons également fait l'hypothèse que MYB221 avait les mêmes cibles.

La ChIP-SEQ a permis d'identifier les cibles potentielles directes de MYB221 à l'échelle du génome. Nos critères de sélection ont permis de retenir 31 cibles à proximité des pics communs de deux expériences ChIP. La moitié des séquences des pics sélectionnés, séquences qui contiennent les sites de fixation du FT, possède un motif SMRE conservé (figII.6, Zhong et Ye 2012). Ce motif est identique au motif MBSIIG par lequel EgMYB1 se fixe spécifiquement aux gènes de biosynthèse de lignines (Legay *et al.* 2007). Dans ces 31 cibles communes aux deux expériences, on ne retrouve pas les gènes *CAD* et *CCR*, alors qu'ils ont été identifiés dans l'une des expériences (*CCR*) ou les deux (*CAD*). De la même façon, on enregistre que *MYB221* est dans la liste des cibles de MYB221 dans les deux expériences, mais avec deux pics différents non superposés. Deux cibles de MYB221 parmi celles testées par qPCR (une O-acyl transférase membranaire et un FT MYB similaire à AtMYB4, connu pour son effet répresseur de la lignification (Jin *et al.* 2000)) sont régulées négativement chez les plantes transgéniques MYB221 OE (figIII.8). Ceci est en accord avec l'hypothèse que MYB221 est un répresseur de transcription. Toutefois, une autre cible testée (alcool déshydrogenase de classe III) est régulée positivement chez ces mêmes plantes transgéniques (figIII.8). Ceci pourrait avoir trois explications possibles : (1) soit cette cible est un faux-positif, ce qui est peu probable car elle possède un motif SMRE, (2) soit elle est, en

même temps, sous le contrôle d'un autre FT inconnu qui serait un activateur, (3) soit sa répression par MYB221 nécessite la présence d'une protéine partenaire, qui n'était pas présente dans l'échantillon biologique considéré.

Les plantes MYB221OE n'ont pas présenté de baisse de lignification (figIII.4-D, E, F ; figIII.6-F) ni de croissance (figIII.3-A, B ; figIII.1-A, B, C) et nous avons évoqué dans le chapitre 3 que cela est certainement lié au faible niveau d'expression de *MYB221* chez les plantes transgéniques (figIII.6-A). En revanche, les plantes répressives dominantes ont montré une baisse considérable de la lignification de leurs parois secondaires, de leur taux de lignines et de leur croissance. Comme pour MYB156, le motif SRDX semble avoir amplifié la répression. Cependant, l'expression des 2 cibles (*MYB4* et *MBOAT*) chez les plantes MYB221-SRDX OE est identique à celle du témoin (figIII.8) alors qu'on pourrait s'attendre à ce qu'elle soit plus réprimée que chez les MYB221 OE. Ceci pourrait être expliqué par un phénomène éventuel de rétrocontrôle de l'expression de ces cibles où un acteur moléculaire pourrait les activer dès que leur niveau d'expression atteint un seuil minimal. Nous supposons ainsi que chez les plantes MYB221 OE, ce seuil minimal n'a pas été atteint car le transgène est faiblement exprimé et n'a engendré qu'une légère répression de l'expression de *MYB4* et *MBOAT*.

MYB4, le gène cible de MYB221, est un facteur MYB appartenant au même sous-groupe des facteurs MYB répresseurs de la lignification. Il possède également le motif répresseur de type EAR. Il n'y a pas de contradiction entre notre hypothèse suggérant que MYB221 est un répresseur de lignification, et le fait que *MYB4* soit un répresseur de gènes de biosynthèse de lignines, mais en même temps réprimé par MYB221. En effet, même si *MYB4* était réprimé chez les transgéniques MYB221-SRDX OE, il pourrait ne pas avoir un rôle central dans la lignification suite à l'implication éventuelle d'autres cascades de régulation y compris, peut-être, une autorégulation négative du gène *MYB221* par le FT MYB221 puisque la séquence ADN de *MYB221* contient le motif SMRE (ACCAACT) dans sa région promotrice et que *MYB221* a été trouvé dans les cibles potentielles de MYB221. Il pourrait exister un contrôle fin de l'expression des gènes de la biosynthèse des lignines *via* des boucles de rétrocontrôle relativement complexe entre MYB221 et MYB4

Pour compléter ces résultats, il serait intéressant de réaliser un nouvel essai ChIP-SEQ en utilisant un matériel végétal plus homogène entre les deux expériences ChIP-SEQ (même période de culture, mêmes conditions de croissance ...) afin d'avoir une nouvelle liste de cibles avec moins de risque de biais. De plus, l'interaction entre MYB221 d'une part et les

promoteurs de *MYB4*, *MYB221* ainsi que d'autres cibles doit être vérifiée par des techniques alternatives *in vitro*.

MYB090 semble être un répresseur direct d'un certain nombre de gènes impliqués dans la formation des parois secondaires *via* un motif similaire à Gamyb

La ChIP-SEQ a permis d'identifier les cibles potentielles de MYB090 à l'échelle du génome, les fonctions des cibles potentielles sont assez diversifiées. Certaines sont des régulateurs transcriptionnels de type MYB, ZFP, bHLH et autres. D'autres sont des enzymes impliquées dans la biosynthèse de polysaccharides, notamment la cellulose (1,3- β -glucane synthase, sucrose synthase, CESA...), et de lignines (CCoAOMT, 4CL). D'autres cibles codent pour des glycosyltransférases notamment GT8 impliquée dans la biosynthèse des xylanes (Zhong *et al.* 2005, Li *et al.* 2011) ainsi que pour des enzymes de modification de la paroi cellulaire (pectine estérase). De plus, plusieurs gènes potentiellement impliqués dans la signalisation hormonale, notamment les auxines (AUX/IAA, auxin influx carrier (AUX1 LAX family)), les cytokinines (cytokinine déshydrogenase) et les gibbérélines (Gibberellin 2-oxidase) ont été identifiés. De même, nous avons trouvé plusieurs cibles potentielles codant pour des protéines à arabinogalactanes à domaine fascicline (FLAs).

Parmi les gènes testés par qPCR chez les plantes transgéniques surexprimant *MYB090*, la majorité semble avoir une expression réprimée (figIII.23). Ainsi, nous postulons que MYB090 est un répresseur transcriptionnel potentiellement impliqué dans la formation de la paroi secondaire.

Dans la majorité des séquences des pics détectés en 5' ou dans le gène, un motif consensus très similaire au motif Gamyb a été retrouvé, en position médiane de la séquence du pic (figII.11). Cela suggère très fortement que MYB090 se fixe sur ces séquences génomiques au niveau de ce motif consensus. Le motif Gamyb correspond au domaine de fixation à l'ADN du FT GAMYB, FT MYB qui appartient au groupe R2R3 et qui est inductible par les gibbérélines. Il joue un rôle dans la signalisation *via* l'acide gibbérélique et la régulation de la plupart des gènes inductibles par les gibbérélines chez le riz (Kaneko *et al.* 2004, Aya *et al.* 2009, Liu *et al.* 2010)).

Les plantes transgéniques MYB090 OE montrent une baisse de leur taux de lignines totales (figIII.21-C). De plus, au niveau histologique, ce sont surtout les parois des rayons

ligneux qui présentent une diminution importante de leur lignification (figIII.19-B ; figIII.20-B). Aussi, bien que certains gènes de biosynthèse de lignines (*PAL1* et *CCR1*) soient régulés chez ces plantes d'une manière positive, un autre gène (*CAD*) est régulé négativement (figIII.22). Parallèlement, on trouve parmi les cibles de MYB090, des gènes de biosynthèse de lignines notamment CCoAOMT et 4CL. Ainsi, nous postulons que MYB090 pourrait être un régulateur négatif de la lignification des rayons ligneux. Pour le vérifier, il est envisageable d'étudier l'expression tissulaire de MYB090 et de gènes de biosynthèse des lignines par hybridation *in situ*.

Nous avons pu identifier aussi des gènes de biosynthèse de cellulose parmi les cibles de MYB090. Ces gènes ne sont pas des gènes de biosynthèse de cellulose impliqués dans la formation de la paroi secondaire (*CESA4*, 7 et 8) (pour revue, voir Mellerowicz et Sundberg 2008). Ils sont probablement impliqués dans la formation de la paroi primaire car ils sont similaires à CESA1 et 6 d'*Arabidopsis* (Desprez *et al.* 2007). Par contre, les deux gènes *CESA4* et *CESA8* sont régulés positivement chez les plantes MYB090 OE (figIII.22). Nous posons l'hypothèse que MYB090 réprime directement l'expression de certains gènes de biosynthèse de cellulose impliqués dans la formation de la paroi primaire. Il régulerait positivement mais de façon indirecte l'expression de ceux ayant un rôle majeur dans la formation de la paroi secondaire. Nous avons essayé de vérifier si le phénotype observé au niveau des rayons ligneux pouvait être dû également à une augmentation de la teneur en cellulose des parois des cellules des rayons. Cependant, les colorations utilisées en études histologiques (Safranine-Bleu Alcian et IKI-H₂SO₄) n'ont pas permis de conclure de façon évidente. La coloration au Calcofluor white par exemple pourrait être envisagée pour vérifier cette hypothèse.

De même, un nombre de glycosyl transférases pouvant être impliquées dans la biosynthèse de la paroi secondaire d'une façon générale, et notamment GT8 impliquée dans la biosynthèse des xylanes (Zhong *et al.* 2005, Li *et al.* 2011), a été identifié comme cibles de MYB090. De plus, le gène *IRX8* est réprimé chez les plantes MYB090 OE. Avec l'anticorps AX1 (qui reconnaît les arabinoxylanes, Guillon *et al.* 2004), on observe une diminution légère de la fluorescence due aux xylanes immunomarqués, ce qui est concordant avec la légère répression de *IRX8*. Toutefois, il serait intéressant de tester d'autres anticorps qui marquent d'autres types de xylanes pour avoir une vue complète de la répartition et de l'abondance de cette famille de polysaccharides.

Plusieurs gènes potentiellement impliqués dans la signalisation avec les régulateurs de croissance notamment les auxines, les cytokinines et les gibbérellines ont été identifiés parmi les cibles. De plus, la croissance des plantes transgéniques MYB090 OE est fortement modifiée au niveau de la morphologie des feuilles, de la longueur des entrenœuds, de la forme de la base de la tige et de la taille des stipules. Ces modifications pourraient avoir une cause hormonale. Par ailleurs, sachant que le matériel utilisé pour la ChIP-SEQ de MYB090 est du xylème tendu (cf Matériel et Méthodes) et que *MYB090* est surexprimé dans ce tissu (figI.2 ; figI.3), nous pouvons postuler que MYB090 pourrait avoir un rôle dans la formation du bois de tension en régulant l'expression des gènes impliqués dans la signalisation avec les gibbérellines car ce régulateur de croissance s'est avéré important pour le développement de la fibre G en longueur et en épaisseur dans le bois de tension chez l'Acacia (Nugroho *et al.* 2013). Des dosages du taux de gibbérellines chez les plantes MYB090 OE par comparaison aux plantes sauvages ainsi que des traitements par des gibbérellines ou des inhibiteurs de gibbérellines pourraient être envisagés pour éclaircir l'éventuel rôle de MYB090 dans cette voie.

Les phénotypes observés chez les plantes transgéniques au niveau de la croissance sont similaires à ceux observés chez des peupliers surexprimant *MIR156* appelé *Corngrass1(Cg1)* sous le contrôle du promoteur 35S. Ces plantes ont montré de la même façon une baisse de leur teneur en lignines (Rubinelli *et al.* 2012). Cette forte ressemblance entre les deux cas peut être expliquée par une éventuelle interaction entre MYB090 et MIR156. Par ailleurs, plusieurs gènes connus pour être régulés par ce micro ARN comme *APETALA1 et 2* et *LEAFY* figurent parmi les cibles directes potentielles de MYB090, ce qui plaide en faveur d'une éventuelle interaction entre MYB090 et MIR156 dans le cadre de la régulation de la croissance des plantes. Une transformation des plantes MYB090 OE ou MYB090-SRDX OE avec une construction permettant l'expression de MIR156 permettrait de mieux comprendre cette interaction dans un premier temps.

Limites des approches ChIP-SEQ

L'expression d'un certain nombre de cibles potentielles de MYB090 ou de MYB221 chez les plantes transgéniques reste inchangée alors qu'on s'attendait à ce qu'elle soit régulée positivement ou négativement. Cependant, il est à noter que dans toutes les études réalisées, il y a une corrélation limitée entre la fixation à l'ADN et la transcription qui en résulte. C'est-à-

dire que certains évènements de fixation à l'ADN n'ont pas d'effet ou seulement un effet faible sur l'expression de gènes. Il peut également exister des redondances fonctionnelles parmi des FT liés. Le FT peut réguler le gène cible dans un autre type de cellules, à un autre stade développemental, ou dans d'autres conditions physiologiques (peut-être en présence d'autres cofacteurs) ou peut dépendre d'un certain évènement de remodelage de la chromatine, sans oublier les artefacts expérimentaux (pour revue voir Ferrier *et al.* 2011). La ChIP fait intervenir des anticorps dont on ne peut être certain à 100% qu'ils ne précipitent que le FT étudié, et donc cela peut générer des faux-positifs.

Si MYB090 aurait un rôle plus ou moins ciblé vers les rayons ligneux et si MYB221 jouerait un rôle plutôt au niveau des fibres, cela expliquerait le fait que la ChIP-SEQ ait moins bien marché sur MYB221 en ce qui concerne le nombre de pics puisqu'il devrait être plus facile de ponter et infiltrer des cellules de rayons. On devrait avoir un biais à ce niveau-là.

La ChIP-SEQ sur MYB221 a donné un nombre de pics communs et donc de cibles très faible par rapport à MYB090. Comme évoqué dans la discussion du chapitre 2, cela peut-être dû à une trop grande variabilité biologique entre les deux échantillons pour MYB221, ou simplement, MYB221 a moins de cibles que MYB090. A la lumière des résultats des transgéniques, MYB090 semble avoir majoritairement une action dans les rayons. Le pontage peut être facilité dans les cellules des rayons dont la paroi est plus fine, et donc la ChIP porte sur une population de cellules de xylème enrichies en cellules de rayons pontées. Au final, la ChIP est plus efficace pour les cibles de MYB090.

Ainsi, afin d'avoir une idée claire sur les fonctions de MYB221 et MYB090, il est indispensable de valider leurs cibles par d'autres techniques alternatives, notamment des analyses d'expression à une échelle plus large en utilisant des puces par exemple. Cette expression serait à tester chez des plantes transgéniques surexprimant les FT mais aussi sur des plantes à différents stades physiologiques, en fonction de la nature de cibles. Ceci aboutirait à un modèle assez complet du réseau régulationnel au sein duquel interviennent les différents FT notamment en ce qui concerne la formation du bois.

En conclusion, les 3 FT étudiés dans notre étude, MYB090, MYB221 et MYB156, semblent être des régulateurs importants de la formation de la paroi secondaire des cellules de xylème chez le peuplier mais aussi de la croissance des plantes, dont les fonctions potentielles

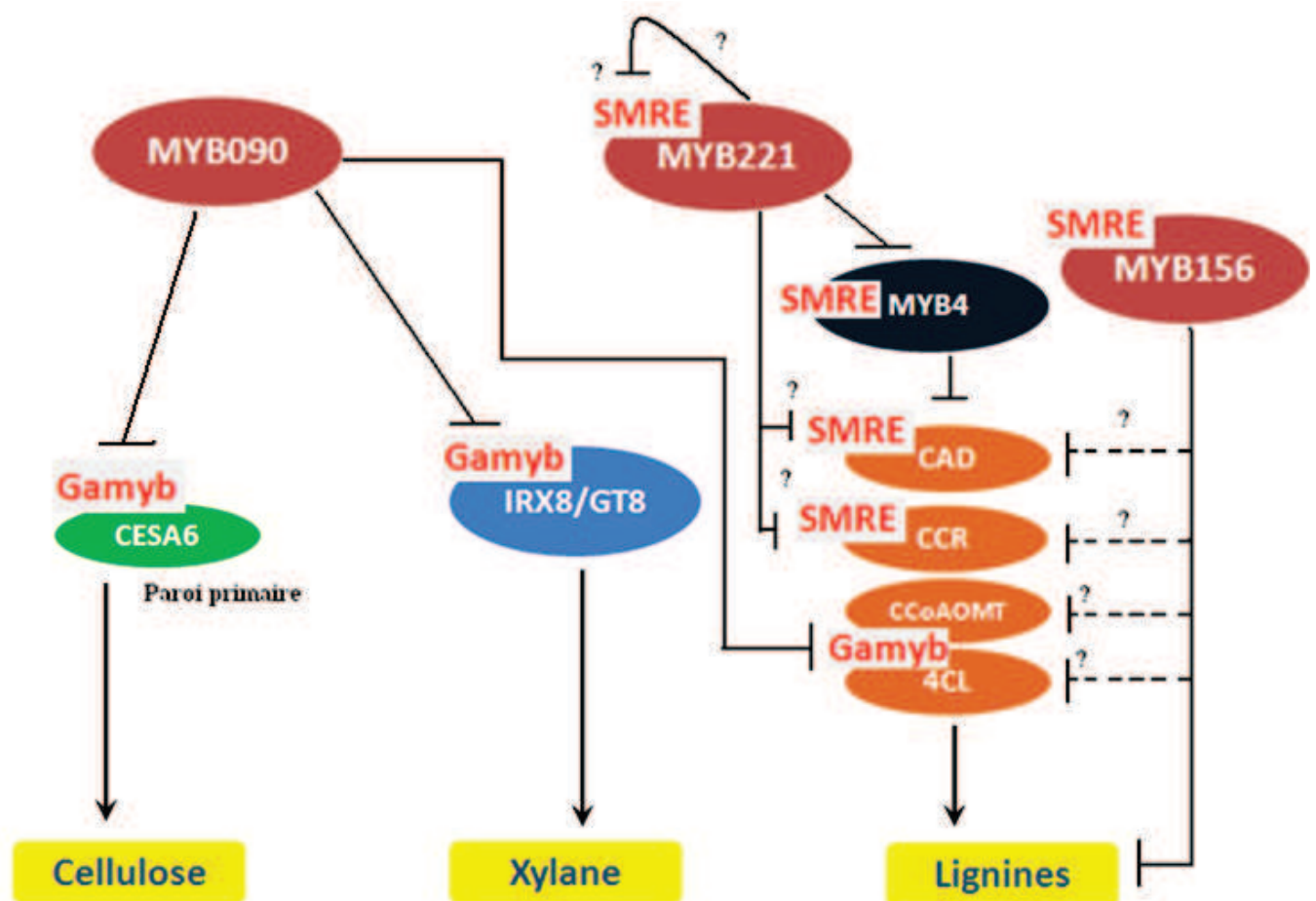


Figure D.1 : Schéma hypothétique représentant les principaux résultats obtenus après notre étude sur MYB090, MYB156 et MYB221.

Les lignes continues représentent des régulations directes potentielles et les lignes discontinues représentent une régulation indirecte potentielle. Les points d'interrogation représentent des hypothèses à vérifier.

sont synthétisées sur la figure D.1. Toutefois, d'autres expériences seraient intéressantes à réaliser afin de mieux comprendre les fonctions précises des différents FT étudiés :

- ChIP-SEQ sur MYB156 et étude de biomasse et de lignification de certaines lignées sur une période de croissance plus longue.
- Nouvelle expérience ChIP-SEQ sur MYB221 avec des échantillons plus homogènes pour minimiser le biais expérimental ainsi qu'une vérification de l'interaction de MYB221 avec les promoteurs de certaines cibles notamment *MYB4* et *MYB221*.
- Etude d'expression de MYB090 au niveau des rayons ligneux par hybridation *in situ* et réalisations d'autres expériences d'histologie notamment la détection de la cellulose avec le Calcofluor white et l'immunomarquage de xylanes à l'aide d'autres anticorps. Les taux de gibbérellines et la réponse à ces phytohormones au niveau des plantes transgéniques peuvent également être étudiés.

Références bibliographiques

Aguayo, M. G., R. T. Mendonca, et al. (2012). "Chemical characteristics and kraft pulping of tension wood from *Eucalyptus globulus* Labill." *Revista Arvore* 36(6): 1163-1171.

Alzari, P. M., H. Berglund, et al. (2006). "Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE." *Acta Crystallographica Section D-Biological Crystallography* 62: 1103-1113.

Ambavaram, M. M. R., A. Krishnan, et al. (2011). "Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice." *Plant Physiology* 155(2): 916-931.

Andersson-Gunnerås, S., E. J. Mellerowicz, et al. (2006). "Biosynthesis of cellulose-enriched tension wood in *Populus*: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis." *Plant Journal* 45(2): 144-165.

Ascenzi R, Gantt JS. (1999). "Subnuclear distribution of the entire complement of linker histone variants in *Arabidopsis thaliana*." *Chromosoma* 108(6): 345-355.

Aya, K., M. Ueguchi-Tanaka, et al. (2009). "Gibberellin modulates anther development in rice *via* the transcriptional regulation of GAMYB." *Plant Cell* 21(5): 1453-1472.

Barski, A. Zhao. K. (2009). "Genomic location analysis by ChIP-Seq." *J Cell Biochem* 107: 11-18.

Berthelot A., C. Bastien, et al. (2005). "Poplar SIG, 4 years after its creation-Le GIS Peuplier, 4 ans après sa creation." *Informations - Foret, Afocel* (708): 6 pp.

Bhargava, A., A. Ahad, et al. (2013). "The interacting MYB75 and KNAT7 transcription factors modulate secondary cell wall deposition both in stems and seed coat in *Arabidopsis*." *Planta* 237(5): 1199-1211.

Boerjan, W., J. Ralph, et al. (2003). "Lignin biosynthesis." *Annual Review of Plant Biology* 54: 519-546.

Bonawitz, N. D. et C. Chapple (2010). "The genetics of lignin biosynthesis: connecting genotype to phenotype". *Annual Review of Genetics*, 44: 337-363.

Bowler, C., G. Benvenuto, et al. (2004). "Chromatin techniques for plant cells." *Plant Journal* 39(5): 776-789.

Bradshaw HD, Jr., Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. *Theoretical and Applied Genetics* 89 (2/3):167-178

Brandt, R., M. Salla-Martret, et al. (2012). "Genome-wide binding-site analysis of REVOLUTA reveals a link between leaf patterning and light-mediated growth responses." *Plant Journal* 72(1): 31-42.

Brown DM, Zeef LA, Ellis J, Goodacre R, Turner SR (2005). "Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics." *Plant Cell* 17: 2281–2295.

Bulger M, Groudine M, (2010). Enhancers : The abundance and function of regulatory sequences beyond promoters. *Developmental Biology* 339(2) : 250-257.

Busov, V. B., A. M. Brunner, et al. (2005). "Genetic transformation: a powerful tool for dissection of adaptive traits in trees." *New Phytologist* 167(1): 9-18.

Cervera MT, Storme. V., Ivens B, Gusmao J, Liu BH, Hostyn V, Slycken J, Montagu Mv, Boerjan W (2001). "Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P.nigra* and *P.trichocarpa*) based on AFLP and microsatellite markers." *Genetics* 158(2): 787-809.

- Chaffey N, Cholewa. E., Regan S, Sundberg B (2002). "Secondary xylem development in Arabidopsis: a model for wood formation." *Physiologia Plantarum* 114(4): 594-600.
- Chang S, Puryear J, et al. (1993). "A simple and efficient method for isolating RNA from pine trees." *Plant Molecular Biology Reporter* 11(2): 113-116.
- Clair B, Almeras T, Yamamoto H, Okuyama T, Sugiyama J, (2006). Mechanical behavior of cellulose microfibrils in tension wood, in relation with maturation stress generation. *Biophysical Journal*, 91:1128–1135.
- Curtis, M. D. et U. Grossniklaus (2003). "A gateway cloning vector set for high-throughput functional analysis of genes *in planta*." *Plant Physiology* 133(2): 462-469.
- Daniel, G., L. Filonova, et al. (2006). "Morphological and chemical characterisation of the G-layer in tension wood fibres of *Populus tremula* and *Betula verrucosa*: Labelling with cellulose-binding module CBM1(HjCel7A) and fluorescence and FE-SEM microscopy." *Holzforschung* 60(6): 618-624.
- Das, P. M., K. Ramachandran, et al. (2004). "Chromatin immunoprecipitation assay." *Biotechniques* 37(6): 961-969.
- De Micco, V., K. Ruel, et al. (2012). "Xylem anatomy and cell wall ultrastructure of *Nicotiana tabacum* after lignin genetic modification through transcriptional activator EgMYB2." *Iawa Journal* 33(3): 269-286.
- Déjardin, A., F. Laurans, et al. (2010). "Wood formation in Angiosperms." *Comptes Rendus Biologies* 333(4): 325-334.
- Déjardin, A., J. C. Leplé, et al. (2004). "Expressed sequence tags from poplar wood tissues - A comparative analysis from multiple libraries." *Plant Biology* 6(1): 55-64.
- Deluc, L., F. Barrieu, et al. (2006). "Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway." *Plant Physiology* 140(2): 499-511.
- Demura, T. et H. Fukuda (2007). "Transcriptional regulation in wood formation." *Trends in Plant Science* 12(2): 64-70.
- Desprez, T., M. Juraniec, et al. (2007). "Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*." *Proceedings of the National Academy of Sciences of the United States of America* 104(39): 15572-15577.
- Dickmann DI, Kuzovkina J. (2008). "Poplars and willows of the world, with emphasis on silviculturally important species " *FAO* (ed).
- Dillen SY, Storme V., Marron N, Bastien C, Neyrinck S, Steenackers M, Ceulemans R, Boerjan W (2009). "Genomic regions involved in productivity of two interspecific poplar families in Europe. 1. Stem height, circumference and volume." *Tree Genetics and Genomes* 5 (1): 147-164.
- Dillen SY, El Kasmoui O, et al. (2011). "Poplar. In: Halford NG KA (ed) *Energy crops*." RSC Publishing, Cambridge: 275-300.
- Donaldson, L. A. (2001). "Lignification and lignin topochemistry - an ultrastructural view." *Phytochemistry* 57(6): 859-873.
- Du, J. and A. Groover (2010). "Transcriptional Regulation of Secondary Growth and Wood Formation." *Journal of Integrative Plant Biology* 52(1): 17-27.

Du Z., Zhou X., Ling Y., Zhang Z, and Su Z. agriGO: a GO analysis toolkit for the agricultural community Nucleic Acids Research Advance Access published on July 1, 2010, DOI 10.1093/nar/gkq310. Nucl. Acids Res. 38: W64-W70.

FCBA. (2009). "Etat de l'art sur les techniques et savoirs faire de récolte des taillis à courtes rotations." www.fcba.fr: pp. 30.

Ferrier, T., J. Tomas Matus, et al. (2011). "Arabidopsis paves the way: genomic and network analyses in crops." Current Opinion in Biotechnology 22(2): 260-270.

Fornale, S., X. H. Shi, et al. (2010). "ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux." Plant Journal 64(4): 633-644.

Fournier M, Baillères H, Chanson B, (1994). Tree biomechanics : growth, cumulative prestresses, and reorientations. Biomimetics, 2: 229-251.

Galis, I., P. Simek, et al. (2006). "A novel R2R3 MYB transcription factor NtMYBJS1 is a methyl jasmonate-dependent regulator of phenylpropanoid-conjugate biosynthesis in tobacco." Plant Journal 46(4): 573-592.

Gaudet M, Jorge V., Paolucci I, Beritognolo I, Mugnozza GS, Sabatti M (2008). "Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait." Tree Genetics & Genomes 4(1): 25-36.

Gierlinger, N. et M. Schwanninger (2006). "Chemical imaging of poplar wood cell walls by confocal Raman microscopy." Plant Physiology 140(4): 1246-1254.

Gilmour DS and Lis JT (1985). "In vivo interactions of RNA polymerase II with genes of *Drosophila melanogaster*." Molecular and Cellular Biology 5(8): 2009-2018.

Goicoechea, M., E. Lacombe, et al. (2005). "EgMYB2, a new transcriptional activator from Eucalyptus xylem, regulates secondary cell wall formation and lignin biosynthesis." Plant Journal 43(4): 553-567.

Grima-Pettenati, J. Soler M., Camxargo, E. L. O., Wang H. (2012). "Transcriptional regulation of the lignin biosynthetic pathway revisited: new players and insights." Advances in Botanical Research 61: 173-218.

Guedes F T P, Laurans F, Quemener B, Assor C, Boizot N, Vigouroux J, Lesage-Descauses M-C, Leplé J-C, Déjardin A, Pilate G. Evaluation of non-cellulosic polysaccharide distribution in differentiating and mature poplar tension wood fibres: abundance of rhamnogalacturonan I and absence of xyloglucan in the G-layer. Soumis à Plant Physiology

Guillaumie, S., R. Mzid, et al. (2010). "The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco." Plant Molecular Biology 72(1-2): 215-234.

Guillon F, Tranquet O, Quillien L, Utile JP, Ortiz JJO, Saulnier L (2004) Generation of polyclonal and monoclonal antibodies against arabinoxylans and their use for immunocytochemical location of arabinoxylans in cell walls of endosperm of wheat. Journal of Cereal Science 40 (2004) 167–182.

Guirimand, G., V. Courdavault, et al. (2010). "Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"?" BMC Plant Biology 10: 182.

Han, X. J., G. He, et al. (2011). "Expression analysis of two NAC transcription factors PtNAC068 and PtNAC154 from Poplar." Plant Molecular Biology Reporter 30(2): 370-378.

- Hertzberg, M., H. Aspeborg, et al. (2001). "A transcriptional roadmap to wood formation." *Proceedings of the National Academy of Sciences of the United States of America* 98(25): 14732-14737.
- Hiratsu, K., K. Matsui, et al. (2003). "Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*." *Plant Journal* 34(5): 733-739.
- Hongkai Ji, H Jiang, et al. (2008). "An integrated software system for analyzing ChIP-chip and ChIP-seq data." *Nature Biotechnology* 26: 1293-1300.
- Hussey S. G., E. Mizrachi, et al. (2011). "SND2, a NAC transcription factor gene, regulates genes involved in secondary cell wall development in *Arabidopsis* fibres and increases fibre cell area in *Eucalyptus*." *BMC Plant Biology* 11: 17.
- Jin, H. L., E. Cominelli, et al. (2000). "Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*." *EMBO Journal* 19(22): 6150-6161.
- Jorge V, Dowkiw A., Faivre-Rampant P, Bastien C (2005). "Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection." *New Phytologist* 167(1): 113-127.
- Joseleau, J. P., T. Imai, et al. (2004). "Detection *in situ* and characterization of lignin in the G-layer of tension wood fibres of *Populus deltoides*." *Planta* 219(2): 338-345.
- Jourez, B., A. Riboux, et al. (2001). "Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* cv 'Ghoy')." *IAWA Journal* 22(2): 133-157.
- Kagale, S. et K. Rozwadowski (2011). "EAR motif-mediated transcriptional repression in plants An underlying mechanism for epigenetic regulation of gene expression." *Epigenetics* 6(2): 141-146.
- Kaneko, M., Y. Inukai, et al. (2004). "Loss-of-function mutations of the rice GAMYB gene impair alpha-amylase expression in aleurone and flower development." *Plant Cell* 16(1): 33-44.
- Karpinska, B., M. Karlsson, et al. (2004). "MYB transcription factors are differentially expressed and regulated during secondary vascular tissue development in hybrid aspen." *Plant Molecular Biology* 56(2): 255-270.
- Kazan, K. (2006). "Negative regulation of defence and stress genes by EAR-motif-containing repressors." *Trends Plant Sci* 11(3): 109-112.
- Kerstens, S., W. F. Decraemer, et al. (2001). "Cell walls at the plant surface behave mechanically like fiber-reinforced composite materials." *Plant Physiology* 127(2): 381-385.
- Ko, J. H., C. Prassinis, et al. (2006). "Developmental and seasonal expression of PtaHB1, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166)." *New Phytologist* 169(3): 469-478.
- Kranz, H. D., M. Denekamp, et al. (1998). "Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*." *Plant Journal* 16(2): 263-276.
- Lee, C., Q. Teng, et al. (2010). "The *Arabidopsis* family GT43 glycosyltransferases form two functionally nonredundant groups essential for the elongation of glucuronoxylan backbone." *Plant Physiology* 153(2): 526-541.
- Legay, S., E. Lacombe, et al. (2007). "Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway." *Plant Science* 173(5): 542-549.

- Legay, S., P. Sivadon, et al. (2010). "EgMYB1, an R2R3 MYB transcription factor from eucalyptus negatively regulates secondary cell wall formation in Arabidopsis and poplar." *New Phytologist* 188(3): 774-786.
- Lehringer, C., N. Gierlinger, et al. (2008). "Topochemical investigation on tension wood fibres of *Acer* spp., *Fagus sylvatica* L. and *Quercus robur* L." *Holzforschung* 62(3): 255-263.
- Leplé, J. C., A. C. M. Brasileiro, et al. (1992). "Transgenic poplars-expression of chimeric genes using 4 different constructs." *Plant Cell Reports* 11(3): 137-141.
- Leplé, J. C., R. Dauwe, et al. (2007). "Downregulation of cinnamoyl-coenzyme a reductase in poplar: Multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure." *Plant Cell* 19(11): 3669-3691.
- Leseberg, C. H., A. L. Li, et al. (2006). "Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*." *Gene* 378: 84-94.
- Letunic, I. and P. Bork (2007). "Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation." *Bioinformatics* 23(1): 127-128.
- Li, E. Y., A. Bhargava, et al. (2012). "The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in Arabidopsis and is functionally conserved in *Populus*." *New Phytologist* 194(1): 102-115.
- Li, Q. Z., Y. C. Lin, et al. (2012). "Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*." *Proceedings of the National Academy of Sciences of the United States of America* 109(36): 14699-14704.
- Li, Q. Z., D. Y. Min, et al. (2011). "Down-regulation of glycosyltransferase 8D genes in *Populus trichocarpa* caused reduced mechanical strength and xylan content in wood." *Tree Physiology* 31(2): 226-236.
- Liu, Z. H., W. J. Bao, et al. (2010). "Identification of gamyb-4 and Analysis of the Regulatory Role of GAMYB in Rice Anther Development." *Journal of Integrative Plant Biology* 52(7): 670-678.
- Ma, Q.-H., C. Wang, et al. (2011). "TaMYB4 cloned from wheat regulates lignin biosynthesis through negatively controlling the transcripts of both cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase genes." *Biochimie* 93(7): 1179-1186.
- Machanick P and Bailey T. L. (2011). "MEME-ChIP: motif analysis of large DNA datasets." *Bioinformatics* 27(12): 1696-1697.
- Marron N, Dillen SY, Ceulemans R (2007) Evaluation of leaf traits for indirect selection of high yielding poplar hybrids. *Environmental and Experimental Botany* 61(2) : 103-116
- McCarthy, R. L., R. Q. Zhong, et al. (2010). "The Poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis." *Plant and Cell Physiology* 51(6): 1084-1090.
- McCarthy, R. L., R. Q. Zhong, et al. (2009). "MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis." *Plant and Cell Physiology* 50(11): 1950-1964.
- Mele, G., N. Ori, et al. (2003). "The knotted1-like homeobox gene BREVIPEDICELLUS regulates cell differentiation by modulating metabolic pathways." *Genes & Development* 17(17): 2088-2093.
- Mellerowicz, E. J., P. Immerzeel, et al. (2008). "Xyloglucan: the molecular muscle of trees." *Annals of Botany* 102(5): 659-665.

- Mellerowicz, E. J. and B. Sundberg (2008). "Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties." *Current Opinion in Plant Biology* 11(3): 293-300.
- Mitsuda, N., A. Iwase, et al. (2007). "NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis." *Plant Cell* 19(1): 270-280.
- Mitsuda, N., M. Seki, et al. (2005). "The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence." *Plant Cell* 17(11): 2993-3006.
- Moreno-Cortes, A., T. Hernandez-Verdeja, et al. (2012). "CsRAV1 induces sylleptic branching in hybrid poplar." *New Phytologist* 194(1): 83-90.
- Mutwil, M., S. Debolt, et al. (2008). "Cellulose synthesis: a complex complex." *Current Opinion in Plant Biology* 11(3): 252-257.
- Nishikubo, N., T. Awano, et al. (2007). "Xyloglucan endo-transglycosylase (XET) functions in gelatinous layers of tension wood fibers in poplar - A glimpse into the mechanism of the balancing act of trees." *Plant and Cell Physiology* 48(6): 843-855.
- Norberg PH, Meier H. (1966). Physical and chemical properties of the gelatinous layer in tension wood fiber of aspen (*Populus tremula* L.). *Holzforschung*, 20: 174–178.
- Nugroho, W. D., S. Nakaba, et al. (2013). "Gibberellin mediates the development of gelatinous fibres in the tension wood of inclined *Acacia mangium* seedlings." *Annals of Botany* 112(7): 1321-1329.
- Ohman, D., B. Demedts, et al. (2013). "MYB103 is required for FERULATE-5-HYDROXYLASE expression and syringyl lignin biosynthesis in Arabidopsis stems." *Plant Journal* 73(1): 63-76.
- Omer, S., S. Kumar, et al. (2013). "Over-expression of a subgroup 4 R2R3 type MYB transcription factor gene from *Leucaena leucocephala* reduces lignin content in transgenic tobacco." *Plant Cell Reports* 32(1): 161-171.
- Paux E., Carocha V., Marques C., de Sousa A.M., Borralho N., Sivadon P., Grima-Pettenati J. (2005). "Transcript profiling of Eucalyptus xylem genes during tension wood formation" *New Phytologist* 167: 89–100.
- Piquemal, J., C. Lapierre, et al. (1998). "Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants." *Plant Journal* 13(1): 71-83.
- Plomion C., G. Leprovost., A. Stokes (2001). "Wood formation in trees." *Plant Physiology* 127: 1513–1523.
- Prashant, S., M. S. Sunita, et al. (2011). "Down-regulation of *Leucaena leucocephala* cinnamoyl CoA reductase (LICCR) gene induces significant changes in phenotype, soluble phenolic pools and lignin in transgenic tobacco." *Plant Cell Reports* 30(12): 2215-2231.
- Preston, J., J. Wheeler, et al. (2004). "AtMYB32 is required for normal pollen development in Arabidopsis thaliana." *Plant Journal* 40(6): 979-995.
- Prouse, M. B. and M. M. Campbell (2012). "The interaction between MYB proteins and their target DNA binding sites." *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* 1819(1): 67-77.
- Qiu, D., I. W. Wilson, et al. (2008). "Gene expression in Eucalyptus branch wood with marked variation in cellulose microfibril orientation and lacking G-layers." *New Phytologist* 179(1): 94-103.
- Redillas, M., J. S. Jeong, et al. (2012). "The overexpression of OsNAC9 alters the root architecture of rice plants enhancing drought resistance and grain yield under field conditions." *Plant Biotechnology Journal* 10(7): 792-805.

- Romano, J. M., C. Dubos, et al. (2012). "AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network." *New Phytologist* 195(4): 774-786.
- Rubinelli P M, G Chuck, et al. (2012). "Constitutive expression of the Corngrass1 microRNA in poplar affects plant architecture and stem lignin content and composition." *Biomass and Bioenergy* 54: 312–321.
- Saito, K., Y. Watanabe, et al. (2012). "Direct mapping of morphological distribution of syringyl and guaiacyl lignin in the xylem of maple by time-of-flight secondary ion mass spectrometry." *Plant Journal* 69(3): 542-552.
- Shen, H., X. Z. He, et al. (2012). "Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks." *New Phytologist* 193(1): 121-136.
- Soler, E., C. Andrieu-Soler, et al. (2011). "A systems approach to analyze transcription factors in mammalian cells." *Methods* 53(2): 151-162.
- Sonbol, F. M., S. Fornale, et al. (2009). "The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*." *Plant Molecular Biology* 70(3): 283-296.
- Stanton B.J. (2009) The domestication and conservation of *Populus* genetic resources (Chapter 4a). In: *Poplars and willows in the world: meeting the needs of society and the environment*. Eds. Isebrands J.G. & Richardson J. FAO/IPC (Food and Agricultural Organization of the United Nations / International Poplar Commission), Rome, Italy. 86p.
- Sterky, F., S. Regan, et al. (1998). "Gene discovery in the wood-forming tissues of poplar: Analysis of 5,692 expressed sequence tags." *Proceedings of the National Academy of Sciences of the United States of America* 95(22): 13330-13335.
- Tamagnone, L., A. Merida, et al. (1998). "The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco." *Plant Cell* 10(2): 135-154.
- Testone, G., E. Condello, et al. (2012). "The peach (*Prunus persica* L. Batsch) genome harbours 10 KNOX genes, which are differentially expressed in stem development, and the class 1 KNOPE1 regulates elongation and lignification during primary growth." *Journal of Experimental Botany* 63(15): 5417-5435.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994 Nov 11;22(22):4673-80.
- Tsai, C. C., L. F. Hung, et al. (2012). "Biomechanical features of eccentric cambial growth and reaction wood formation in broadleaf tree branches." *Trees-Structure and Function* 26(5): 1585-1595.
- Tuskan, G. A., S. DiFazio, et al. (2006). "The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)." *Science* 313(5793): 1596-1604.
- Valouev, A., D. S. Johnson, et al. (2008). "Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data." *Nature Methods* 5(9): 829-834.
- Vanholme, R., K. Morreel, et al. (2008). "Lignin engineering." *Current Opinion in Plant Biology* 11(3): 278-285.
- Wang, H. Z., U. Avci, et al. (2010). "Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants." *Proceedings of the National Academy of Sciences of the United States of America* 107(51): 22338-22343.

- Wilbanks, E. G., D. J. Larsen, et al. (2012). "A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq." *Nucleic Acids Research* 40(10): 11.
- Wilkins, O., H. Nahal, et al. (2009). "Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors." *Plant Physiology* 149(2): 981-993.
- Xu, B., N. Sathitsuksanoh, et al. (2012). "Overexpression of AtLOV1 in Switchgrass alters plant architecture, lignin content, and flowering time." *Plos One* 7(12): 11.
- Yant, L. (2012). "Genome-wide mapping of transcription factor binding reveals developmental process integration and a fresh look at evolutionary dynamics." *American Journal of Botany* 99(2): 277-290.
- Ye, Z. H., W. S. York, et al. (2006). "Important new players in secondary wall synthesis." *Trends in Plant Science* 11(4): 162-164.
- Yordanov, Y. S., S. Regan, et al. (2010). "Members of the LATERAL ORGAN BOUNDARIES DOMAIN transcription factor family are involved in the regulation of secondary growth in *Populus*." *Plant Cell* 22(11): 3662-3677.
- Zhong, R., E. A. Richardson, et al. (2007c). "The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*." *Plant Cell* 19(9): 2776-2792.
- Zhong, R. a. Y., Z-H. (2009). "Transcriptional regulation of lignin biosynthesis." *Plant Signaling Behavior* 4(11): 1028-1034.
- Zhong, R. Q., T. Demura, et al. (2006). "SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*." *Plant Cell* 18(11): 3158-3170.
- Zhong, R. Q., C. H. Lee, et al. (2010a). "Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis." *Trends in Plant Science* 15(11): 625-632.
- Zhong, R. Q., C. H. Lee, et al. (2010b). "Functional characterization of Poplar wood-associated NAC domain transcription factors." *Plant Physiology* 152(2): 1044-1055.
- Zhong, R. Q., C. H. Lee, et al. (2008). "A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*." *Plant Cell* 20(10): 2763-2782.
- Zhong, R. Q., R. L. McCarthy, et al. (2011). "Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in Poplar." *Plant Physiology* 157(3): 1452-1468.
- Zhong, R. Q., E. A. Richardson, et al. (2007b). "Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*." *Planta* 225(6): 1603-1611.
- Zhong, R. Q. and Z. H. Ye (2007a). "Regulation of cell wall biosynthesis." *Current Opinion in Plant Biology* 10(6): 564-572.
- Zhong, R. Q. and Z. H. Ye (2012). "MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes." *Plant and Cell Physiology* 53(2): 368-380.
- Zhong, R. Q., M. J. Pena, et al. (2005). "*Arabidopsis* fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis." *Plant Cell* 17(12): 3390-3408.
- Zhou, J. L., C. H. Lee, et al. (2009). "MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*." *Plant Cell* 21(1): 248-266.

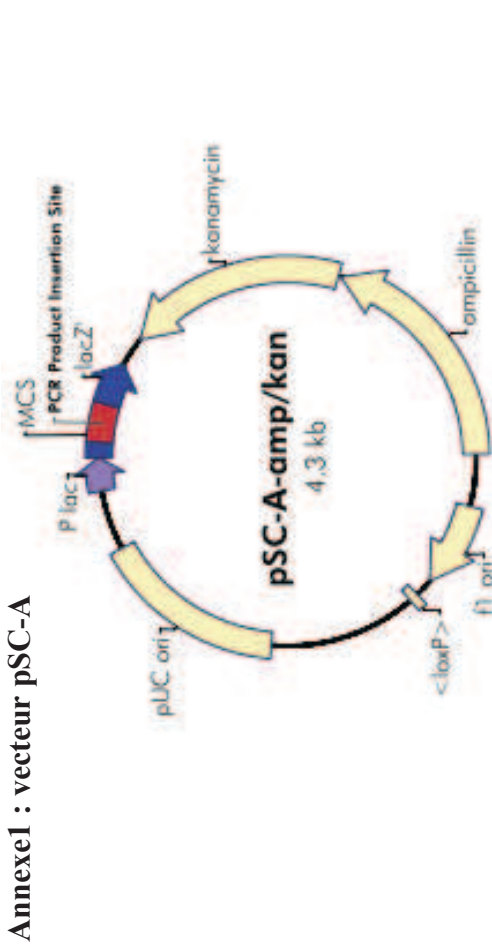
Zhu L, Shan. H., Chen S, Jiang J, Gu C, et al. (2013). "The heterologous expression of the Chrysanthemum R2R3-MYB transcription factor CmMYB1 alters lignin composition and represses flavonoid synthesis in *Arabidopsis thaliana*." PLoS ONE 8(6): e65680.

Zimmermann I. M., Heim M. A., Weisshaar B. and Uhrig J. F. Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. The Plant Journal (2004) 40, 22–34.

Zinzen, R. P., C. Girardot, et al. (2009). "Combinatorial binding predicts spatio-temporal cis-regulatory activity." Nature 462(7269): 65-70.

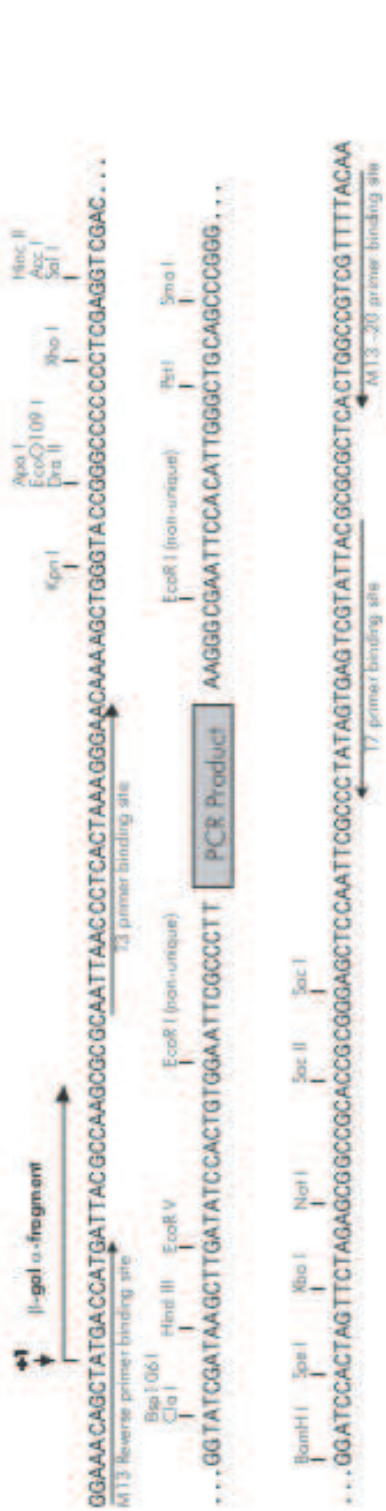
Annexes

Annexel : vecteur pSC-A



pSC-A-amp/kan PCR Cloning Vector PCR Product Insertion Site Region

(sequence shown 4261–4270, 1–250)



Note The pSC-A-amp/kan plasmid shown represents the end product of the StrataClone™ PCR cloning system. It is derived from topoisomerase I-mediated ligation of two linear StrataClone vector arms to a PCR product of interest, followed by Cre-mediated recombination. Please see the StrataClone™ PCR Cloning Kit instruction manual for additional information.

Annexe 3 : vecteur pDONOR Zéo (Invitrogen)

Map and Features of pDONOR™221 and pDONOR™/Zeo

Map of pDONOR™221 and pDONOR™/Zeo

The map below shows the elements of pDONOR™221 and pDONOR™/Zeo. The complete sequences of pDONOR™221 and pDONOR™/Zeo are available for downloading from our Web site (www.invitrogen.com) or by contacting Technical Support (page 20).



Comments for:

- m13 T2 transcription termination sequence (ct):
- m13 T1 transcription termination sequence (ct):
- M13 Forward (-20) priming site:
- amp^r:
- ccdB gene (ct):
- Chloramphenicol resistance gene (ct):
- amp^r2 (ct):
- T7 Promoter/priming site (ct):
- M13 Reverse priming site:
- Kanamycin resistance gene:
- EM7 promoter (ct):
- Zeocin resistance gene (cp):
- pUC origin:

Recombination Region of pDONOR™221 and pDONOR™/Zeo

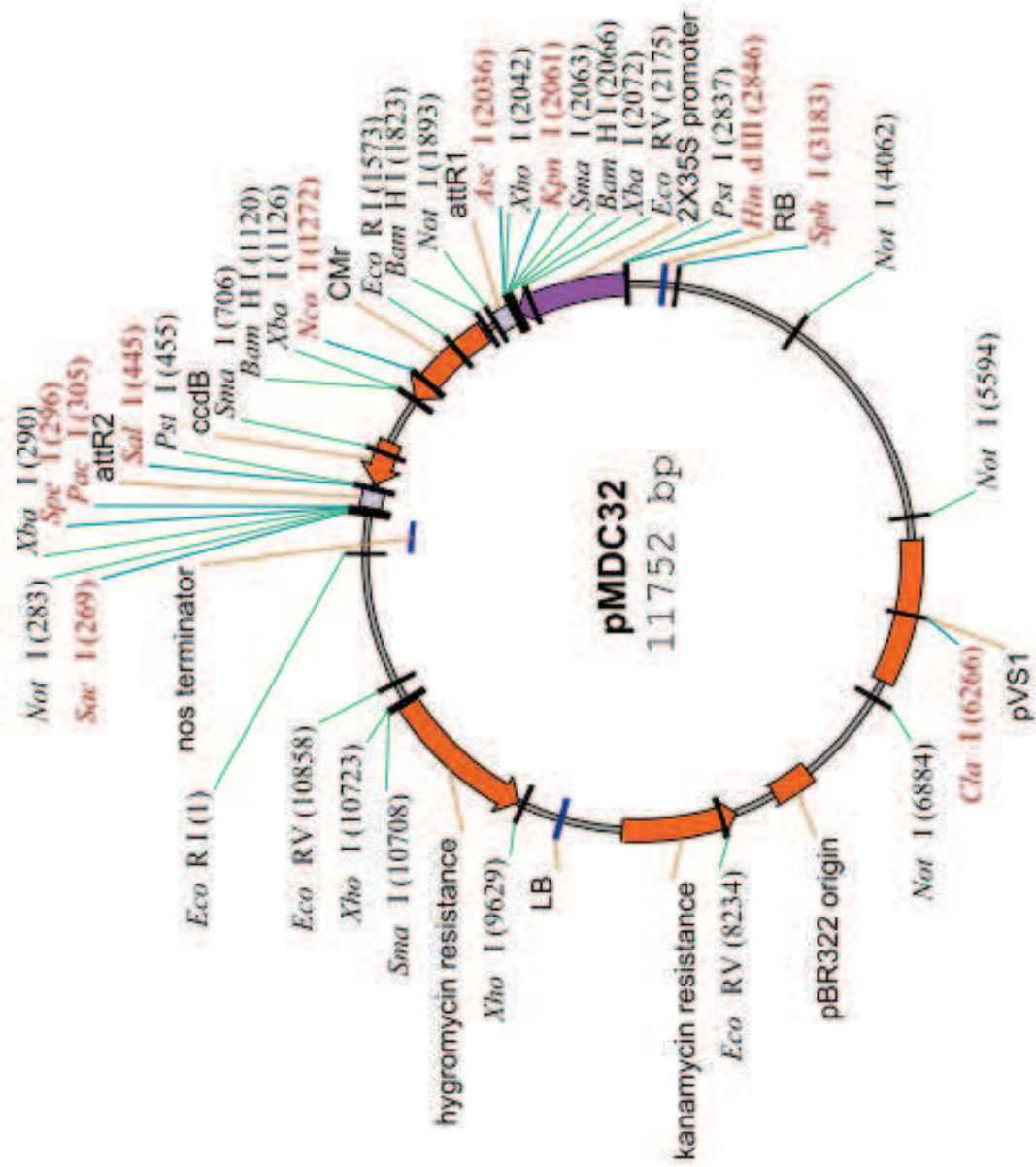
The recombination region of the expression clone resulting from pDONOR™221 x entry clone or pDONOR™/Zeo x entry clone is shown below.

Features of the Recombination Region:

- Shaded regions correspond to DNA sequences transferred from the attB substrate into pDONOR™221 or pDONOR™/Zeo by recombination. Non-shaded regions are derived from the pDONOR™221 or pDONOR™/Zeo vector.
- Bases 651 and 2897 of the pDONOR™221 or pDONOR™/Zeo vector sequence are marked.



Annexe 4 : vecteur pMDC32



Annexe 5 : Criblage des lignées transgéniques

nom de la transformation	n° de la lignée	Niveau d'expression du transgène	Criblage niveau 1	fibres - lignifiées	rayons - lignifiés	taille + grande	taille + petite	diamètre + petit	diamètre + grand	beaucoup de syléptiques	racines + grandes	feuilles anormales	stipule + grand	Criblage niveau 2	Criblage niveau 3
MYB221	1	xx	Choisie			x		x		x				Choisie	Choisie
MYB221	8	xx	Choisie						x	x				Choisie	Choisie
MYB221	11	xx	Choisie			x			x	xx				Choisie	Choisie
MYB221	12	xx	Choisie			x				x				Choisie	Choisie
MYB221	18	xx	Choisie						x	xxx				Choisie	Choisie
MYB221	7		tt							xx				tt	tt
MYB221	9	xx	Choisie			x			x	x				Choisie	
MYB221	20	xx	Choisie				x			xxx				Choisie	
MYB221	5	xx	Choisie				x	x		x				Non	
MYB221	16	xx	Choisie							xx				Non	
MYB221	22	xx	Choisie			x								Non	
MYB221	2		Non							x					
MYB221	3		Non							x					
MYB221	4	x	Non							x					
MYB221	6	x	Non							x					
MYB221	10		Non							xxx					
MYB221	14	x	Non			x				xxxx					
MYB221	15	x	Non												
MYB221	17		Non						x	xxx					
MYB221	19	x	Non												
MYB221	21	x	Non					x							
MYB221	24	x	Non				x								
MYB221-SRDX	2	xx	Choisie	xxxx	x		xxx	xxx		x		x		Choisie	Choisie
MYB221-SRDX	8	xx	Choisie	xx	x		x	x		x		x		Choisie	Choisie
MYB221-SRDX	17	xx	Choisie	xx	x		xx	xx		x				Choisie	Choisie
MYB221-SRDX	18		tt											tt	tt
MYB221-SRDX	7	xx	Choisie	xxx	x			x		xx				Choisie	
MYB221-SRDX	10	xx	Choisie	x		x				xx				Non	
MYB221-SRDX	6	xx	Choisie	~		x		x		xx				Non	
MYB221-SRDX	14	xx	Choisie					x						Non	
MYB221-SRDX	19	xx	Choisie	x	x					x				Non	
MYB221-SRDX	20	xx	Choisie			x				x					
MYB221-SRDX	21	xx	Choisie					x		x					
MYB221-SRDX	22	xx	Choisie							x					
MYB221-SRDX	24	xx	Choisie							x					
MYB221-SRDX	1	x	Non					x		x		x			
MYB221-SRDX	3	x	Non				xxx	xxx				xx			
MYB221-SRDX	4	x	Non			x				xx					

[illegible]

Annexe 6 : Rapport du projet de séquençage de l'ADN Immunoprécipité pour MYB221

Project: S13117

This report presents an overview of the quality controls performed on samples from the S13117 project sequenced on the HISO70 flowcell. A detailed report is also given for each samples in separate pdf files.

Contents

- 1. Number of sequenced reads 1
- 2. Base quality 1
- 3. Size and MD5 2

1. Number of sequenced reads

Table 1 provides some sequencing statistics. The platform provides a minimum of 25 millions reads per sample for standard projects.

Sample ID	Sample name	Raw reads	Aligned reads	% aligned	Unique positions	% unique
ADN-03.B4	10174B IP/MYB1 170713	50,582,208	29,037,263	57.41	7,139,604	24.59
ADN-04.B5	10174B Input	56,311,227	35,564,542	63.16	28,640,002	80.53

Table 1: Sequencing statistics for each sample.
Values for which a quality threshold has been set up by the platform are colored. In the "Raw reads" column, if the number of reads is equal or above 25 million then cells are colored in green, otherwise they are colored in red. The "Aligned reads" column represents the number of reads aligned onto the reference genome. "%aligned" is the ratio (aligned reads / total reads). The "Unique positions" column contains the number of different positions in the genome to which reads are aligned. "%unique" is the ratio (unique positions / aligned reads). The aligned reads and unique position numbers are for information purpose only.

2. Base quality

Table 2 shows the percentage of read positions with an average base quality over 30. The average base quality of 95% of the positions of the reads must be over 30 to pass the platform quality criteria. A base quality of 30 means that there is 1 chance out of 1000 that this base has been miscalled during the Base Calling step (i.e. probability of error = 0.001).

Figure 1 shows the average base quality at each position along the read. The quality threshold defined by the platform (average base quality = 30) is highlighted with a red line.

Sample ID	Sample name	Average base quality over 30
ADN-03.B4	10174B IP/MYB1 170713	100.00%
ADN-04.B5	10174B Input	100.00%

Table 2: Percentage of read positions with an average base quality over 30.
Values for which a quality threshold has been set up by the platform are colored. If the average base qualities are above 30 for 95% of the read position, then corresponding cell is colored in green, otherwise it is colored in red.

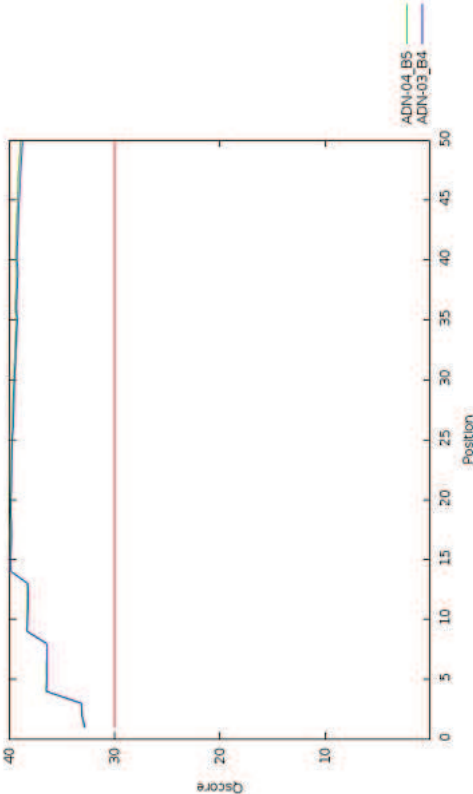


Figure 1: Average base quality at each position along the read.

3. Size and MD5

Total size of the project's data: 7.1 Gb
The size of each file is given in Table 3.

File	Size (Mb)
ADN-03.B4.fastq.gz	1,884
ADN-03.B4.sort.bam	991
ADN-03.B4.sort.bam.bai	1
ADN-03.B4.sort.bed.gz	323
ADN-04.B5.fastq.gz	2,074
ADN-04.B5.sort.bam	1,401
ADN-04.B5.sort.bam.bai	1
ADN-04.B5.sort.bed.gz	481
wigs_for_ADN-03.B4.wig.gz	41
wigs_for_ADN-04.B5.wig.gz	44

Table 3: Size of the project's files.

You can use MD5 strings to check the integrity of your files after the download. They are provided in the file md5.txt. To get more information about MD5 usage, please go to our dedicated wiki section : http://ags-igbac.fr/wiki/doku.php?id=md5_usage

Project: S13136

Annexe 7 : Rapport du projet de séquençage de l'ADN Immunoprécipité pour MYB221

This report presents an overview of the quality controls performed on samples from the S13136 project sequenced on the Illumina HiSeq. A detailed report is also given for each sample in separate pdf files.

Contents	
1. Number of sequenced reads	1
2. Base quality	1
3. Size and MD5	2

1. Number of sequenced reads

Table 1 provides some sequencing statistics. The platform provides a minimum of 25 millions reads per sample for standard projects.

Sample ID	Sample name	Raw reads	Aligned reads	% aligned	Unique positions	% unique
ADN-05p.B1	10174C IP/MYB21	59,635,214	29,418,903	49.33	7,442,705	25.30
ADN-06.B2	10174D IP/MYB21	62,039,805	32,672,386	52.66	2,754,123	8.43
ADN-07.B5	10174D Input	53,192,248	33,764,646	63.48	19,797,416	58.63

Table 1: Sequencing statistics for each sample.
Values for which a quality threshold has been set up by the platform are colored. In the "Raw reads" column, if the number of reads is equal or above 25 million then cells are colored in green, otherwise they are colored in red. The "Aligned reads" column represents the number of reads aligned onto the reference genome. "%aligned" is the ratio (aligned reads / total reads). The "Unique positions" column contains the number of different positions in the genome to which reads are aligned. "%unique" is the ratio (unique positions / aligned reads). The aligned reads and unique position numbers are for information purpose only.

2. Base quality

Table 2 shows the percentage of read positions with an average base quality over 30. The average base quality of 95% of the positions of the reads must be over 30 to pass the platform quality criteria. A base quality of 30 means that there is 1 chance out of 1000 that this base has been miscalled during the Base Calling step (i.e, probability of error = 0.001).

Figure 1 shows the average base quality at each position along the read. The quality threshold defined by the platform (average base quality = 30) is highlighted with a red line.

Sample ID	Sample name	Average base quality over 30
ADN-05p.B1	10174C IP/MYB21	100.00%
ADN-06.B2	10174D IP/MYB21	100.00%
ADN-07.B5	10174D Input	100.00%

Table 2: Percentage of read positions with an average base quality over 30.
Values for which a quality threshold has been set up by the platform are colored. If the average base qualities are above 30 for 95% of the read position, then corresponding cell is colored in green, otherwise it is colored in red.

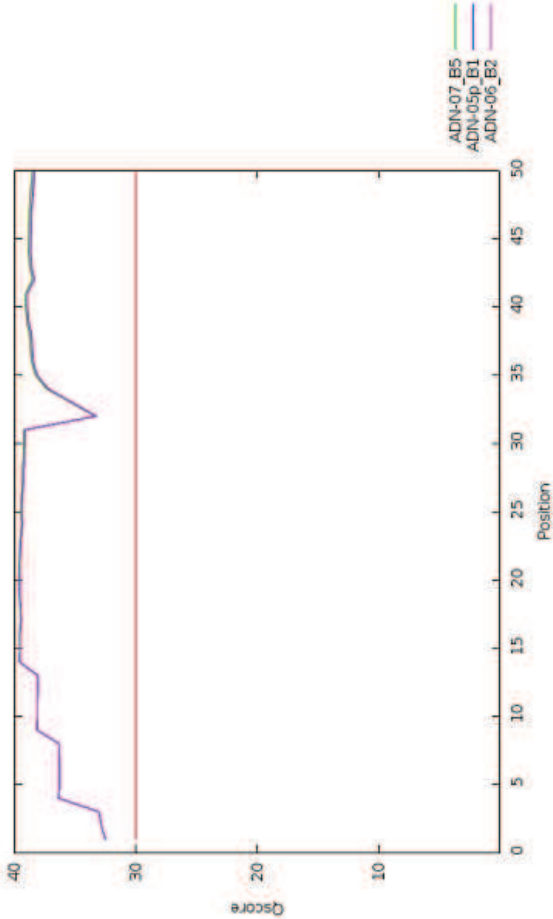


Figure 1: Average base quality at each position along the read.

3. Size and MD5

Total size of the project's data: 11.4 Gb
The size of each file is given in Table 3.

File	Size (Mb)
------	-----------

Table 3: Size of the project's files.

You can use MD5 strings to check the integrity of your files after the download. They are provided in the file md5.txt. To get more information about MD5 usage, please go to our dedicated wiki section : http://ags.igbmc.fr/wiki/doku.php?id=md5_usage

Annexe 8 : Liste de cibles de MYB221 identifiées par ChIP-SEQ

Numéros des Pics communs	Chromosome	Start_Min	Stop_Max	Position du pic par rapport au gene	Modèle du gène cible	Description de la fonction du gène cible
148-6	scaffold_25	88450	88849	in_gene	Potri.T011400	Transcription factor, Myb superfamily
198-429	Chr02	23488800	23489099	in_gene	Potri.002G242500	Cytochrome b5-like Heme/Steroid binding domain
217-1525	scaffold_394	10150	10349	in_gene	Potri.T143600	lipid metabolism
309-33	Chr13	1969300	1969599	in_gene	Potri.013G029800	SHIKIMATE DEHYDROGENASE
368-454	Chr04	5599200	5599449	in_gene	Potri.004G067000	Alcohol dehydrogenase, class III
421-203	Chr13	686000	686249	in_gene	Potri.013G010900	protein-nucleus import
473-139	scaffold_578	20450	20649	in_gene	Potri.T153200	pas d'annotation
539-435	Chr04	4225850	4226149	in_gene	Potri.004G053800	lipid metabolism
560-483	Chr01	42398400	42398599	in_gene	Potri.001G402800	Regulator of chromosome condensation (RCC1) repeat
769-438	Chr08	9656300	9656549	in_gene	Potri.008G143600	tetratricopeptide repeat tpr like superfamily protein
834-27	Chr11	15979550	15979849	in_gene	Potri.011G135900	Glucose-6-phosphate/phosphate and phosphoenolpyruvate/phosphate antiporter
12-2	Chr19	409850	410399	5'	Potri.019G004000	Protein of unknown function DUF88
249-22	Chr05	8767700	8768049	5'	Potri.005G114000	FE(II)/ ASCORBATE OXIDASE SUPERFAMILY
249-22	Chr05	8767700	8768049	5'	Potri.005G114100	pas d'annotation
25-14	Chr13	36450	36849	5'	Potri.013G000600	membrane bound O-acyl transferase (MBOAT) family protein; similar to low similarity to skinny hedgehog (Drosophila melanogaster) Gl:15420842; [ortholog of At1g57600,]
308-176	Chr11	4554800	4555049	5'	Potri.011G053000	Protein of unknown function (DUF616)
351-11	Chr18	10971050	10971399	5'	Potri.018G082900	fatty acid metabolism
353-714	Chr06	25747150	25747449	5'	Potri.006G251800	bZIP transcription factor
402-172	Chr11	11549000	11549249	5'	Potri.011G095000	pas d'annotation
410-200	Chr08	3319050	3319349	5'	Potri.008G056100	Protein of unknown function (DUF506)
422-42	Chr08	14196600	14197049	5'	Potri.008G201600	pas d'annotation
492-161	Chr14	2684550	2684849	5'	Potri.014G032800	hydrolase
512-30	Chr01	8165900	8166199	5'	Potri.001G102200	pas d'annotation
660-1703	Chr08	8442400	8442549	5'	Potri.008G129200	hydroxyproline-rich glycoprotein family protein; similar to proline-rich extensin domain containing protein
682-369	scaffold_20	29300	29649	5'	Potri.T000300	terpene synthase like
728-672	Chr04	1289200	1289399	5'	Potri.004G019400	Glycosyl hydrolase family 1
742-1106	Chr11	3733150	3733299	5'	Potri.011G044000	Cotton fibre expressed protein
76-38	Chr10	15020450	15020749	5'	Potri.010G137000	UDP-GLUCURONOSYLTRANSFERASE RELATED
769-438	Chr08	9656300	9656549	5'	Potri.008G143800	pas bon en fait Potri.008G143600

816-28	Chr11	16691750	16692049	5'	Potri.011G147500	TRANSMEMBRANE PROTEIN INDUCED BY TUMOR NECROSIS FACTOR ALPHA
827-53	Chr02	5525250	5525549	5'	Potri.002G079500	AMINO ACID TRANSPORTER
869-867	Chr04	19461850	19461999	5'	Potri.004G176400	Protein of unknown function (DUF1624)Heparan-α-glucosaminide N-acetyltransferase
869-867	Chr04	19461850	19461999	5'	Potri.004G176500	ER to Golgi transport
87-218	Chr03	14874600	14874899	5'	Potri.003G128600	aquaporin PIP
101-193	Chr02	24133350	24133649	3'	Potri.002G251800	similar to expressed protein in Arabidopsis thaliana; [ortholog of At4g13220.]
184-201	scaffold_25	93250	93799	3'	Potri.T011400	DNA binding
184-201	scaffold_25	93250	93799	3'	Potri.T011500	
309-33	Chr13	1969300	1969599	3'	Potri.013G029900	shikimate 5-dehydrogenase activity
371-138	Chr08	4185700	4185999	3'	Potri.008G068800	NLI interacting factor (NIF) family protein
411-1507	Chr07	14765700	14765849	3'	Potri.007G134500	similar to putative MYB transcription factor
422-42	Chr08	14196600	14197049	3'	Potri.008G201500	similar to Ribonucleoside-diphosphate reductase small chain
427-1736	Chr02	23529150	23529349	3'	Potri.002G242900	
434-1582	Chr03	18769750	18769999	3'	Potri.003G180500	
434-1582	Chr03	18769750	18769999	3'	Potri.003G180600	
492-161	Chr14	2684550	2684849	3'	Potri.014G032900	protein binding
503-1414	Chr15	14075650	14075799	3'	Potri.015G129200	similar to INCOMPLETE ROOT HAIR ELONGATION
512-30	Chr01	8165900	8166199	3'	Potri.001G102100	peptidase M1 family protein
539-435	Chr04	4225850	4226149	3'	Potri.004G053900	protein dimerization activity
550-1247	Chr15	8348450	8348599	3'	Potri.015G060900	
727-60	Chr14	5541650	5541999	3'	Potri.014G068100	similar to DNA repair protein RAD51 homolog 3 (DNA repair-recombination protein RAD51C) (AtRAD51C)
728-672	Chr04	1289200	1289399	3'	Potri.004G019500	hydrolase activity, hydrolyzing O-glycosyl compounds
769-438	Chr08	9656300	9656549	3'	Potri.008G143700	
79-10	Chr05	3186900	3187249	3'	Potri.005G043300	3-dehydroquinase dehydratase activity
79-10	Chr05	3186900	3187249	3'	Potri.005G043400	shikimate 5-dehydrogenase activity
816-28	Chr11	16691750	16692049	3'	Potri.011G147600	dehydration-responsive family protein; similar to early-responsive to dehydration stress ERD3 protein (Arabidopsis thaliana)
868-1171	Chr11	15531800	15531949	3'	Potri.011G129800	protein binding
87-218	Chr03	14874600	14874899	3'	Potri.003G128500	similar to protease-associated (PA) domain-containing protein
11-84	Chr01	16601100	16601499			
28-3	Chr03	7871600	7872049			
6-26	Chr09	10824300	10824699			
192-590	Chr19	11308500	11308699			
211-592	Chr03	17013950	17014199			

23-1446	scaffold_57	230050	230399				
282-29	Chr01	7102700	7102999				
322-689	Chr10	13560650	13560899				
352-45	Chr02	19459750	19460049				
364-776	Chr06	15346600	15346799				
40-70	Chr07	933450	933799				
408-246	Chr17	5450700	5451199				
413-151	Chr04	19273650	19273949				
433-155	Chr15	12340000	12340249				
461-52	Chr11	7438900	7439249				
47-186	Chr06	1030850	1031099				
49-596	Chr17	13769000	13769249				
562-246	Chr17	5450700	5451199				
582-148	Chr01	1601750	1601999				
671-87	Chr13	4671450	4671849				
786-792	Chr06	25939050	25939299				
804-88	Chr02	9165000	9165249				
851-368	Chr06	21291600	21291799				
910-202	Chr12	11192400	11192699				

Surligné en jaune les gènes cibles retenus pour les analyses ChIP-PCR (expression de gène et enrichissement)

Annexe 9 : Liste de cibles de MYB090 identifiées par ChIP-SEQ

Numéros des pics communs	Chromosome	Start_Min	Stop_Max	Position du pic par rapport au gène	Modèle du gène cible	Description de la fonction du gène cible
1-1	Chr19	12880000	12880549	in_gene	Potri.019G098900	protein binding
2-3	Chr03	20368600	20369099	in_gene	Potri.003G201700	
12-15	Chr02	10465000	10465549	in_gene	Potri.002G141000	
1001-574	Chr02	12003350	12003699	in_gene	Potri.002G159800	
1007-646	Chr15	10700200	10700699	in_gene	Potri.015G082700	DNA binding
1010-819	Chr06	21896850	21897099	in_gene	Potri.006G203500	protein kinase activity
101-241	scaffold_49	240600	241099	in_gene	Potri.T054700	DNA binding
1020-354	Chr06	4781350	4781599	in_gene	Potri.006G065100	phospholipid binding
102-209	Chr16	12396300	12396749	in_gene	Potri.016G119100	
1022-878	Chr05	6016000	6016349	in_gene	Potri.005G081700	
103-66	Chr06	1739100	1739499	in_gene	Potri.006G025300	GTPase activity GTP binding
106-113	Chr13	15546200	15546799	in_gene	Potri.013G151300	
107-214	Chr16	11776000	11776449	in_gene	Potri.016G113600	
1077-544	Chr01	5396300	5396599	in_gene	Potri.001G067700	histone binding
1096-586	Chr05	19974600	19974899	in_gene	Potri.005G183100	
1107-889	Chr02	3665500	3665699	in_gene	Potri.002G055000	regulation of transcription, DNA-dependent nucleus DNA binding response to hormone stimulus
1134-795	Chr06	10734350	10734599	in_gene	Potri.006G130900	GTPase activity GTP binding
1146-770	Chr04	16042950	16043199	in_gene	Potri.004G138500	
115-191	Chr16	14430150	14430549	in_gene	Potri.016G143700	
116-176	Chr10	19203600	19204149	in_gene	Potri.010G199900	transferase activity
1167-764	Chr01	28395450	28395749	in_gene	Potri.001G277300	iron ion binding electron carrier activity oxidoreductase activity
1207-912	Chr10	16394650	16394949	in_gene	Potri.010G156900	heme binding
1224-857	scaffold_1264	7550	7899	in_gene	Potri.T173600	
1232-590	Chr17	1479250	1479599	in_gene	Potri.017G017600	DNA binding
125-329	Chr06	5226550	5226949	in_gene	Potri.006G069800	
1316-745	Chr14	6124050	6124399	in_gene	Potri.014G075600	protein binding protein binding
1319-799	Chr06	11491150	11491499	in_gene	Potri.006G137900	
1331-908	Chr13	1888550	1888899	in_gene	Potri.013G028300	signal transducer activity two-component signal transduction system (phosphorelay)
133-271	Chr13	14495950	14496399	in_gene	Potri.013G134300	
1394-413	scaffold_694	5750	6149	in_gene	Potri.T158300	
1395-709	Chr12	15405950	15406399	in_gene	Potri.012G139800	transcription factor activity zinc ion binding
149-139	Chr19	14280450	14280799	in_gene	Potri.019G114800	protein binding
1535-558	Chr02	24789200	24789499	in_gene	Potri.002G259300	ADP binding protein binding signal transduction intracellular ATP binding
1546-991	Chr05	7771100	7771449	in_gene	Potri.005G101600	protein binding
155-122	Chr09	11111300	11111899	in_gene	Potri.009G138900	transmembrane transport
156-127	Chr09	5770300	5770799	in_gene	Potri.009G053900	DNA binding
158-282	Chr10	18141800	18142149	in_gene	Potri.010G184300	mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase activity cytoplasm
159-89	Chr06	7420300	7420899	in_gene	Potri.006G097100	oxidoreductase activity
16-18	Chr03	10953300	10953799	in_gene	Potri.003G081500	triacylglycerol lipase activity lipid metabolism
162-83	Chr10	11510100	11510599	in_gene	Potri.010G091800	
163-174	Chr19	14893250	14893649	in_gene	Potri.019G121100	

164-172	Chr01	4189450	4189899	in_gene	Potri.001G055300	
166-921	Chr02	8965900	8966149	in_gene	Potri.002G119400	
1712-859	Chr04	16033000	16033299	in_gene	Potri.004G138400	ferrochelatase activity
171-302	Chr06	9049750	9050549	in_gene	Potri.006G115300	
1717-551	Chr08	6888500	6888849	in_gene	Potri.008G108200	
172-294	Chr15	12453050	12453499	in_gene	Potri.015G106900	DNA binding regulation of transcription
174-75	Chr08	14348400	14348699	in_gene	Potri.008G202200	ATP binding RNA processing
175-175	Chr01	48338550	48338999	in_gene	Potri.001G449100	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
178-177	Chr06	1283000	1283499	in_gene	Potri.006G019200	metal ion binding metal ion transport metal ion binding metal ion transport
1817-964	Chr05	10018800	10019199	in_gene	Potri.015G075300	protein binding
182-192	Chr10	18409500	18409899	in_gene	Potri.010G188100	protein binding
1823-716	Chr18	16221800	16222099	in_gene	Potri.018G142100	protein binding
1930-980	Chr17	2626000	2626249	in_gene	Potri.017G029900	protein kinase activity ATP binding protein amino acid phosphorylation
194-219	Chr10	14056200	14056649	in_gene	Potri.010G124200	
195-134	Chr19	12612400	12612799	in_gene	Potri.019G095300	
201-195	Chr03	17296350	17296799	in_gene	Potri.003G162000	membrane
203-138	Chr19	12860400	12860749	in_gene	Potri.019G098700	protein binding ADP binding ATP binding apoptosis
204-133	Chr11	17031700	17032699	in_gene	Potri.011G151900	
2049-790	Chr10	18029500	18029749	in_gene	Potri.010G182800	transferase activity, transferring hexosyl groups metabolism transferase activity
2084-188	Chr11	15858450	15858849	in_gene	Potri.011G134200	protein binding
2092-891	Chr08	3998400	3998649	in_gene	Potri.008G066000	
212-236	Chr04	9283750	9284149	in_gene	Potri.004G105500	
214-242	Chr13	685900	686349	in_gene	Potri.013G010900	protein binding protein transporter activity protein-nucleus import nucleus cytoplasm
2161-842	Chr19	13908100	13908449	in_gene	Potri.019G109900	protein binding ATP binding protein amino acid phosphorylation protein kinase activity
220-189	Chr05	24547200	24547799	in_gene	Potri.005G238600	ATP binding phosphotransferase activity, alcohol group as acceptor carbohydrate metabolism
221-768	Chr07	3997000	3997349	in_gene	Potri.007G045200	
226-517	Chr08	10829800	10830249	in_gene	Potri.008G159500	
230-211	Chr16	924300	924649	in_gene	Potri.016G017200	transferase activity, transferring hexosyl groups metabolism transferase activity, transferring hexosyl groups metabolism
2319-867	Chr03	16692350	16692549	in_gene	Potri.003G153600	
2335-259	Chr06	10014500	10014699	in_gene	Potri.006G124800	metal ion binding metal ion transport metal ion binding metal ion transport metal ion binding metal ion transport metal ion binding metal ion transport
235-270	Chr05	24191550	24191999	in_gene	Potri.005G234100	hydrogen ion transporter activity ATP hydrolysis coupled proton transport
2363-718	Chr16	3665850	3666099	in_gene	Potri.016G055700	
237-700	Chr01	9498800	9499249	in_gene	Potri.001G118300	copper ion binding amine oxidase activity quinone binding
240-438	Chr10	7811000	7811399	in_gene	Potri.010G046800	signal transducer activity response to light signal transducer activity
242-201	Chr08	6930400	6930799	in_gene	Potri.008G109100	heat shock protein binding
246-155	Chr06	7557150	7557699	in_gene	Potri.006G098300	
247-200	Chr08	7956600	7957299	in_gene	Potri.008G122100	DNA binding
2494-552	Chr09	12035050	12035299	in_gene	Potri.009G154000	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
249-74	Chr19	14153750	14154199	in_gene	Potri.019G112800	protein binding ADP binding signal transduction intracellular
250-247	Chr12	10457800	10458199	in_gene	Potri.012G078700	oxidoreductase activity metabolism oxidation reduction oxidoreductase activity
2527-844	Chr09	9339800	9340099	in_gene	Potri.009G107600	oxidoreductase activity
2535-833	Chr03	17529200	17529449	in_gene	Potri.003G164800	
254-412	Chr10	19749750	19750149	in_gene	Potri.010G207400	protein binding
2559-877	Chr01	13574150	13574449	in_gene	Potri.001G162400	pectinesterase activity cell wall modification cell wall enzyme inhibitor activity
266-349	Chr06	27636450	27636949	in_gene	Potri.006G276700	
271-400	Chr13	933150	933649	in_gene	Potri.013G014700	hydrolase activity
2807-502	Chr07	1953450	1953699	in_gene	Potri.007G026200	
28-31	Chr01	8442800	8443299	in_gene	Potri.001G106200	actin binding intracellular actin binding intracellular actin binding intracellular
287-861	Chr09	3605350	3605699	in_gene	Potri.009G024700	
293-256	Chr04	7118950	7119349	in_gene	Potri.004G085600	hydrolase activity

296-664	Chr03	20082600	20083199	in_gene	Potri.003G198200	phosphate metabolism
301-437	Chr14	8242100	8242749	in_gene	Potri.014G104800	DNA binding regulation of transcription, DNA-dependent DNA binding regulation of transcription, DNA-dependent
3047-911	Chr19	14062100	14062199	in_gene	Potri.019G111500	
306-708	Chr03	17396400	17396749	in_gene	Potri.003G163000	
319-324	Chr06	23159500	23159899	in_gene	Potri.006G219300	protein kinase activity ATP binding
323-305	Chr14	3691450	3691899	in_gene	Potri.014G047400	
328-251	Chr06	24316650	24317049	in_gene	Potri.006G233400	acetylglucosaminyltransferase activity
329-136	Chr01	5610400	5610899	in_gene	Potri.001G070400	protein binding
331-869	Chr10	18330050	18330499	in_gene	Potri.010G187100	nucleotide binding metal ion binding
337-685	Chr03	14672550	14672899	in_gene	Potri.003G125700	metalloendopeptidase activity
344-599	Chr08	1479600	1479949	in_gene	Potri.008G027600	
349-660	Chr14	10844000	10844299	in_gene	Potri.014G142400	
352-267	Chr01	10755300	10755749	in_gene	Potri.001G133800	membrane
354-809	Chr10	14524050	14524449	in_gene	Potri.010G130100	proteolysis and peptidolysis aspartic-type endopeptidase activity
356-533	Chr03	10045050	10045449	in_gene	Potri.003G071800	serine-type endopeptidase activity proteolysis and peptidolysis
365-368	Chr01	19579900	19580349	in_gene	Potri.001G200500	protein binding
375-132	Chr13	214050	214449	in_gene	Potri.013G003200	
379-671	Chr16	14184350	14184799	in_gene	Potri.016G139700	acid phosphatase activity
380-356	Chr01	10593950	10594249	in_gene	Potri.001G131500	
381-295	Chr16	6807100	6807449	in_gene	Potri.016G086300	protein binding
394-449	Chr12	134000	134499	in_gene	Potri.012G001000	catalytic activity trehalose biosynthesis
396-365	Chr06	23343350	23343749	in_gene	Potri.006G221000	UDP-N-acetylmutamate dehydrogenase activity
397-385	Chr04	1430850	1431399	in_gene	Potri.004G021500	protein binding
407-312	Chr06	22533500	22533799	in_gene	Potri.006G210900	
414-345	Chr13	8382200	8382649	in_gene	Potri.013G087100	protein binding
417-581	Chr01	30745600	30745949	in_gene	Potri.001G304200	
4245-286	Chr04	17655900	17657199	in_gene	Potri.004G155300	nucleic acid binding
425-322	Chr10	14567400	14567849	in_gene	Potri.010G130600	
438-310	scaffold_268	41750	42099	in_gene	Potri.T133400	protein binding
447-555	Chr01	11323500	11324199	in_gene	Potri.001G140700	3-beta-hydroxy-delta5-steroid dehydrogenase activity
448-482	Chr03	9420750	9421149	in_gene	Potri.003G066000	phosphoric diester hydrolase activity metal ion binding guanosine tetraphosphate (5'-ppGpp-3')
450-713	Chr08	13527350	13527749	in_gene	Potri.008G193900	
451-441	Chr02	4075750	4076049	in_gene	Potri.002G059900	protein kinase activity ATP binding protein amino acid phosphorylation
453-399	Chr17	876850	877199	in_gene	Potri.017G010200	protein kinase activity ATP binding protein amino acid phosphorylation
455-473	Chr01	50018550	50019049	in_gene	Potri.001G467000	protein kinase activity ATP binding protein amino acid phosphorylation
459-588	Chr02	5168200	5168649	in_gene	Potri.002G074800	amino acid binding metabolism amino acid binding metabolism
461-512	Chr02	14848000	14848599	in_gene	Potri.002G188700	DNA binding nucleus DNA binding nucleus DNA binding nucleus
46-78	Chr02	1387550	1388049	in_gene	Potri.002G023000	
476-451	Chr13	13958950	13959349	in_gene	Potri.013G127000	protein binding
477-542	Chr10	15511550	15511949	in_gene	Potri.010G144600	
483-775	Chr14	2147600	2147949	in_gene	Potri.014G025200	transcription factor activity
484-663	Chr02	7610800	7611149	in_gene	Potri.002G104900	
485-825	Chr08	11666300	11666649	in_gene	Potri.008G170700	
4875-286	Chr04	17655900	17657199	in_gene	Potri.004G155300	nucleic acid binding
496-897	Chr06	11205500	11206049	in_gene	Potri.006G135600	regulation of transcription, DNA-dependent transcription regulator activity
501-202	Chr05	937900	938299	in_gene	Potri.005G012100	protein binding
503-397	Chr05	444450	444699	in_gene	Potri.005G007600	protein binding
50-43	Chr02	12434850	12435399	in_gene	Potri.002G164700	
505-624	Chr07	8656200	8656549	in_gene	Potri.007G066800	protein binding
517-170	Chr07	15039800	15040299	in_gene	Potri.007G139100	
517-170	Chr07	15039800	15040299	in_gene	Potri.007G139200	protein binding

525-268	Chr04	351750	352099	in_gene	Potri.004G005500	malonyl-CoA decarboxylase activity fatty acid biosynthesis
536-597	Chr08	10212000	10212399	in_gene	Potri.008G151000	copper ion binding electron carrier activity
5376-442	Chr05	9629400	9629549	in_gene	Potri.005G123700	
541-277	Chr09	11894850	11895199	in_gene	Potri.009G151400	[metal ion transporter activity metal ion transport
546-578	Chr11	17186050	17186549	in_gene	Potri.011G153300	DNA binding regulation of transcription, DNA-dependent
54-85	Chr07	13259650	13260099	in_gene	Potri.007G109100	DNA binding regulation of transcription, DNA-dependent
552-410	Chr13	15549700	15550199	in_gene	Potri.013G151400	protein binding
5565-429	Chr11	15859250	15859499	in_gene	Potri.011G134200	nucleic acid binding
571-702	Chr09	5552600	5553249	in_gene	Potri.009G050700	methionine adenosyltransferase activity
5829-946	Chr08	6235300	6235499	in_gene	Potri.008G099300	
587-925	Chr10	10016250	10016599	in_gene	Potri.010G073800	
593-230	Chr17	8182950	8183249	in_gene	Potri.017G071200	ATP binding kinase activity metabolism ATP binding kinase activity metabolism
604-997	Chr08	5283350	5283749	in_gene	Potri.008G084100	oxidoreductase activity metabolism oxidoreductase activity
6129-662	Chr02	7116350	7116699	in_gene	Potri.002G098600	
615-394	Chr01	23960200	23960749	in_gene	Potri.001G227400	hydrolase activity, hydrolyzing O-glycosyl compounds
6221-286	Chr04	17655900	17657199	in_gene	Potri.004G155300	nucleic acid binding
62-40	Chr05	848450	848799	in_gene	Potri.005G011600	protein binding
625-916	Chr11	2589600	2589999	in_gene	Potri.011G031000	regulation of transcription, DNA-dependent
636-951	Chr09	8780500	8780899	in_gene	Potri.009G098100	cysteine-type peptidase activity proteolysis and peptidolysis
643-821	Chr15	870600	870949	in_gene	Potri.015G013300	
663-863	Chr04	16507350	16507749	in_gene	Potri.004G141700	
669-687	Chr04	8173150	8173449	in_gene	Potri.004G095100	zinc ion binding nucleic acid binding zinc ion binding nucleic acid binding
672-523	Chr01	43492550	43492949	in_gene	Potri.001G410700	FMN binding oxidoreductase activity electron carrier activity
675-495	Chr19	15129100	15129499	in_gene	Potri.019G123100	
679-535	Chr19	1402300	1402699	in_gene	Potri.019G011700	
685-589	Chr06	4446650	4446949	in_gene	Potri.006G060800	
686-553	Chr14	12335750	12336149	in_gene	Potri.014G158200	
687-422	Chr02	15746800	15747199	in_gene	Potri.002G196600	
698-602	Chr10	20208800	20209149	in_gene	Potri.010G214600	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds nitrogen metabolism
699-649	Chr03	14238700	14239049	in_gene	Potri.003G119800	
703-793	Chr06	9430400	9430749	in_gene	Potri.006G119300	serine carboxypeptidase activity proteolysis and peptidolysis
707-577	Chr08	5838900	5839349	in_gene	Potri.008G093100	
715-756	Chr02	16450400	16450749	in_gene	Potri.002G202500	pectinesterase activity cell wall modification cell wall
716-836	Chr16	1281300	1281799	in_gene	Potri.016G022900	
720-797	Chr04	12021700	12022099	in_gene	Potri.004G124400	
722-828	Chr06	3890000	3890399	in_gene	Potri.006G053500	transcription factor activity regulation of transcription, DNA-dependent
727-665	Chr11	14068650	14068999	in_gene	Potri.011G115600	transcription factor activity regulation of transcription, DNA-dependent
729-830	Chr10	17708300	17708699	in_gene	Potri.010G177700	zinc ion binding oxidoreductase activity oxidation reduction
731-404	Chr02	414150	414399	in_gene	Potri.002G007700	protein binding
73-45	Chr18	1220750	1221299	in_gene	Potri.018G015100	aspartic-type endopeptidase activity proteolysis and peptidolysis
746-762	Chr02	13278200	13278649	in_gene	Potri.002G174000	phosphate binding photosynthesis protein stabilization photosystem II membrane
746-762	Chr02	13278200	13278649	in_gene	Potri.002G174100	protein binding
750-285	Chr10	17076500	17076899	in_gene	Potri.010G168200	oxidoreductase activity, acting on CH-OH group of donors FAD binding
77-199	Chr16	9893950	9894499	in_gene	Potri.016G101200	transferase activity, transferring glycosyl groups membrane
775-619	Chr01	20338250	20338899	in_gene	Potri.001G204600	protein binding
793-987	Chr09	9031850	9032299	in_gene	Potri.009G102700	oxidoreductase activity oxidation reduction copper ion binding
79-87	Chr15	11707000	11707549	in_gene	Potri.015G096900	
803-629	Chr09	4447350	4447649	in_gene	Potri.009G034500	copper ion binding oxidation reduction oxidoreductase activity
805-655	Chr02	14374950	14375299	in_gene	Potri.002G184000	
817-503	Chr09	4385950	4386249	in_gene	Potri.009G033900	iron ion binding electron carrier activity oxidoreductase activity
825-417	Chr14	3167050	3167349	in_gene	Potri.014G038600	protein kinase activity ATP binding protein amino acid phosphorylation

826-917	Chr11	561150	561449	in_gene	Potri.011G007500	
827-841	Chr07	10078450	10078749	in_gene	Potri.007G076800	transcription factor activity regulation of transcription, DNA-dependent
835-516	Chr10	14677150	14677499	in_gene	Potri.010G131900	ATP binding ATPase activity ATP binding ATPase activity ATP binding
838-622	Chr05	7410650	7410949	in_gene	Potri.005G097800	
842-712	Chr01	44558500	44558899	in_gene	Potri.001G420200	membranetransmembrane transport membranetransmembrane transport
848-874	Chr09	10744500	10744849	in_gene	Potri.009G132500	
852-781	Chr12	4622800	4623199	in_gene	Potri.012G049000	
853-254	Chr01	8370400	8370699	in_gene	Potri.001G104900	
854-858	Chr07	1187800	1188099	in_gene	Potri.007G016100	
855-554	scaffold_321	46250	46549	in_gene	Potri.T139900	
860-526	Chr02	15172800	15173049	in_gene	Potri.002G191400	GTP binding small GTPase mediated signal transduction signal transducer activity
887-543	Chr01	25189100	25189449	in_gene	Potri.001G240900	
900-601	Chr09	12532400	12532799	in_gene	Potri.009G163700	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
90-216	Chr10	21452750	21453199	in_gene	Potri.010G233800	
911-711	Chr05	8526350	8526699	in_gene	Potri.005G110800	structural constituent of ribosome protein biosynthesis ribosome
919-540	Chr10	17539300	17539599	in_gene	Potri.010G174900	protein binding
92-33	Chr19	14227200	14227499	in_gene	Potri.019G113700	ADP binding ATP binding apoptosis ADP binding ATP binding apoptosis
93-102	Chr03	21438300	21438649	in_gene	Potri.003G218400	protein binding
94-109	Chr02	10341800	10342249	in_gene	Potri.002G139400	
948-977	Chr04	16069400	16069849	in_gene	Potri.004G138700	zinc ion binding oxidoreductase activity oxidation reduction zinc ion binding
95-72	Chr16	9583050	9583499	in_gene	Potri.016G098400	protein phosphatase type 2A, intrinsic regulator activity signal transduction
971-628	Chr04	3257950	3258299	in_gene	Potri.004G042800	
97-185	Chr01	14581050	14581499	in_gene	Potri.001G172100	
989-678	Chr07	1103400	1103749	in_gene	Potri.007G014700	protein kinase activity ATP binding protein amino acid phosphorylation
996-612	Chr02	2712300	2712649	in_gene	Potri.002G042200	iron ion binding electron carrier activity oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen heme binding oxidation reduction
3-2	Chr01	42928500	42928999	5'	Potri.001G406900	
5-6	Chr02	13069100	13069599	5'	Potri.002G172100	transcription regulator activity, similar to bHLH protein family
6-7	Chr18	2316550	2317049	5'	Potri.018G029400	cellulose synthase (UDP-forming) activity cellulose biosynthesis
1002-549	Chr14	4698550	4698899	5'	Potri.014G060700	
100-86	Chr12	3991050	3991499	5'	Potri.012G044000	
1022-878	Chr05	6016000	6016349	5'	Potri.005G081800	
1022-878	Chr05	6016000	6016349	5'	Potri.005G081900	
1023-986	Chr11	17450200	17450499	5'	Potri.011G156700	
103-66	Chr06	1739100	1739499	5'	Potri.006G025200	
104-128	Chr10	17792600	17793099	5'	Potri.010G178800	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
104-128	Chr10	17792600	17793099	5'	Potri.010G178900	
104-128	Chr10	17792600	17793099	5'	Potri.010G179100	
1045-962	Chr08	9508150	9508549	5'	Potri.008G142400	
1049-955	Chr05	8594100	8594449	5'	Potri.005G112000	DNA binding
105-135	Chr02	7322700	7323149	5'	Potri.002G101300	copper ion binding electron carrier activity copper ion binding electron carrier activity
1063-903	Chr03	12025000	12025299	5'	Potri.003G092900	
1093-600	Chr05	1876050	1876299	5'	Potri.005G025300	actin binding actin cytoskeleton organization and biogenesis F-actin capping protein complex
1096-586	Chr05	19974600	19974899	5'	Potri.005G183200	
1110-470	Chr04	24009850	24010149	5'	Potri.004G234100	
1134-795	Chr06	10734350	10734599	5'	Potri.006G131000	galactosyl galactosyl xylosyl protein 3-beta-glucuronosyltransferase
115-191	Chr16	14430150	14430549	5'	Potri.016G143600	nucleotidyltransferase activity biosynthesis nucleotidyltransferase activity biosynthesis nucleotidyltransferase
1155-486	Chr08	5668450	5668699	5'	Potri.008G090400	integral to membrane
1161-949	Chr07	10599150	10599499	5'	Potri.007G081900	protein binding
1176-483	Chr02	7499200	7499499	5'	Potri.002G103500	nucleus
1182-519	Chr13	13106900	13107199	5'	Potri.013G118000	
119-160	Chr12	12856550	12857049	5'	Potri.012G104400	

1207-912	Chr10	16394650	16394949	5'	Potri.010G157000	nucleic acid binding
1209-930	Chr10	18445100	18445349	5'	Potri.010G188500	protein binding
122-306	Chr10	13174400	13174849	5'	Potri.010G113300	
1224-857	scaffold_1264	7550	7899	5'	Potri.T173400	oxidoreductase activity[metabolism]oxidoreductase activity[metabolism]
124-120	Chr11	17340400	17340849	5'	Potri.011G155100	protein binding[nucleus]DNA binding[protein binding]nucleus
1249-591	Chr19	3613550	3613799	5'	Potri.019G031600	DNA binding[regulation of transcription, DNA-dependent]
125-329	Chr06	5226550	5226949	5'	Potri.006G069900	
1256-807	Chr02	10803600	10803999	5'	Potri.002G145000	
1266-881	Chr05	22920650	22921049	5'	Potri.005G216200	RNA binding
127-59	Chr10	12842600	12842849	5'	Potri.010G109100	
1281-943	Chr15	12986650	12986949	5'	Potri.015G114300	copper ion binding[electron carrier activity]
1281-943	Chr15	12986650	12986949	5'	Potri.015G114400	
128-69	Chr10	18104300	18104799	5'	Potri.010G183900	defense response[response to biotic stimulus]
129-154	Chr19	14894150	14894599	5'	Potri.019G121100	
1294-988	Chr12	390600	390949	5'	Potri.012G004500	
1299-968	Chr14	5368250	5368549	5'	Potri.014G066200	
1304-755	Chr01	35837800	35838149	5'	Potri.001G351200	transporter activity[oligopeptide transport]
131-205	Chr04	5421750	5422199	5'	Potri.004G065000	oxidoreductase activity
1328-592	Chr10	4531350	4531649	5'	Potri.010G030900	
1331-908	Chr13	1888550	1888899	5'	Potri.013G028400	
1350-469	Chr03	18258600	18258899	5'	Potri.003G173300	
135-196	Chr06	25753000	25753499	5'	Potri.006G251900	cellulose synthase (UDP-forming) activity
1354-610	Chr14	1366050	1366299	5'	Potri.014G013200	
136-140	Chr05	19000400	19000849	5'	Potri.005G174600	membrane
1368-895	Chr05	23842650	23842949	5'	Potri.005G229500	cellular metabolism[oxidoreductase activity]
1383-696	Chr12	2726400	2726649	5'	Potri.012G030500	
1392-896	Chr14	3431950	3432249	5'	Potri.014G042500	
139-318	Chr05	21115850	21116399	5'	Potri.005G194100	
1396-952	Chr01	11907400	11907699	5'	Potri.001G146200	DNA binding[two-component response regulator activity]
140-104	Chr04	3932100	3932549	5'	Potri.004G050500	O-methyltransferase activity[methyltransferase activity]
141-340	Chr12	12837000	12837449	5'	Potri.012G104200	
1425-792	Chr01	13078900	13079299	5'	Potri.001G157500	metalloendopeptidase activity[zinc ion binding]
143-60	Chr13	12657400	12657799	5'	Potri.013G113100	DNA binding[regulation of transcription, DNA-dependent]
144-484	Chr14	9677800	9678249	5'	Potri.014G125700	microtubule motor activity[ATP binding]
1498-634	Chr10	12722200	12722549	5'	Potri.010G107500	oxidoreductase activity
1500-854	Chr08	12096700	12097049	5'	Potri.008G177100	ubiquitin-protein ligase activity
150-103	Chr09	10792700	10793199	5'	Potri.009G133400	serine-type endopeptidase activity
15-16	Chr02	1201750	1202199	5'	Potri.002G020200	
1522-914	Chr14	9446150	9446649	5'	Potri.014G122100	
1526-659	Chr02	11163300	11163549	5'	Potri.002G149900	
157-111	Chr03	9832850	9833199	5'	Potri.003G070200	
160-344	Chr17	4162050	4162449	5'	Potri.017G049800	
1627-931	Chr15	13770200	13770499	5'	Potri.015G124200	
166-276	Chr05	24058600	24059049	5'	Potri.005G232300	UDP-N-acetylmuramate dehydrogenase activity
167-193	Chr01	5838000	5838399	5'	Potri.001G073600	protein kinase activity[ATP binding]
1718-887	Chr10	18876700	18876999	5'	Potri.010G195000	DNA binding
1737-676	Chr09	12014600	12014849	5'	Potri.009G153700	
1737-676	Chr09	12014600	12014849	5'	Potri.009G153800	electron carrier activity[iron-sulfur cluster binding]
176-224	Chr10	20738300	20739399	5'	Potri.010G223200	regulation of transcription, DNA-dependent[nucleus]
176-243	Chr10	20738300	20739399	5'	Potri.010G223200	
1823-716	Chr18	16221800	16222099	5'	Potri.018G142200	
183-52	Chr03	16857950	16858299	5'	Potri.003G156100	peroxidase activity[heme binding]response to oxidative stress[oxidation reduction]

184-130	Chr10	12652450	12652899	5'	Potri.010G106700	protein binding
186-361	Chr18	14942850	14943299	5'	Potri.018G127300	protein kinase activity ATP binding protein amino acid phosphorylation
1878-919	Chr02	23472300	23472599	5'	Potri.002G242200	
189-727	Chr03	18003850	18004349	5'	Potri.003G170000	DNA binding protein dimerization activity transcription factor activity
190-88	Chr10	15418300	15418699	5'	Potri.010G142800	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
191-309	Chr04	5111000	5111399	5'	Potri.004G007900	intracellular protein transport endoplasmic reticulum integral to membrane
19-17	Chr14	12058200	12058849	5'	Potri.014G155000	protein amino acid phosphorylation ATP binding protein kinase activity
192-341	Chr02	9217850	9218249	5'	Potri.002G123300	transcription factor activity sequence-specific DNA binding
196-108	Chr05	24045950	24046349	5'	Potri.005G232200	methyltransferase activity metabolism methyltransferase activity
1990-791	Chr04	15915450	15915749	5'	Potri.004G137400	
1990-791	Chr04	15915450	15915749	5'	Potri.004G137500	
199-147	scaffold_61	64700	65249	5'	Potri.T059900	
202-197	Chr08	6663450	6663849	5'	Potri.008G105100	protein kinase activity ATP binding protein amino acid phosphorylation
2022-848	Chr03	14001800	14002049	5'	Potri.003G116600	amino acid binding metabolism amino acid binding metabolism
20-27	Chr03	9882500	9883099	5'	Potri.003G070500	
205-141	Chr13	14288050	14288549	5'	Potri.013G131600	integral to membrane
2078-909	Chr08	11579100	11579349	5'	Potri.008G169400	enzyme inhibitor activity pectinesterase activity enzyme inhibitor activity pectinesterase activity
210-222	Chr08	8773100	8773449	5'	Potri.008G132600	response to aluminum ion
216-321	Chr09	2954850	2955299	5'	Potri.009G017900	inotropic glutamate receptor activity glutamate-gated ion channel activity
218-720	Chr05	25499150	25499649	5'	Potri.005G253800	protein binding
219-524	Chr06	6336700	6337149	5'	Potri.006G083800	
220-189	Chr05	24547200	24547799	5'	Potri.005G238500	
2201-969	Chr09	11344350	11344749	5'	Potri.009G142900	
22-20	Chr03	15937400	15938149	5'	Potri.003G142800	
2224-414	Chr01	7211100	721249	5'	Potri.001G011400	peroxidase activity heme binding response to oxidative stress oxidation reduction
2241-339	Chr01	49641200	49641399	5'	Potri.001G462300	oxidoreductase activity FAD binding oxidation reduction oxidoreductase activity FAD binding oxidation reduction
227-300	Chr08	8225950	8226449	5'	Potri.008G126100	phosphatidylserine biosynthesis
2281-984	Chr18	10810200	10810449	5'	Potri.018G081000	ATP binding protein binding ATP binding protein binding
2281-984	Chr18	10810200	10810449	5'	Potri.018G081100	transport transmembrane transport membrane amino acid transmembrane transport
228-376	Chr06	19708000	19708399	5'	Potri.006G183200	serine carboxypeptidase activity proteolysis and peptidolysis serine carboxypeptidase activity
229-314	Chr06	12440750	12441099	5'	Potri.006G146000	
231-181	scaffold_94	106300	106799	5'	Potri.T083800	
232-319	Chr01	25778800	25779299	5'	Potri.001G247600	membrane transmembrane transport membrane transmembrane transport
23-36	Chr12	559350	559749	5'	Potri.012G006400	O-methyltransferase activity methyltransferase activity protein dimerization activity
234-369	Chr09	7220000	7220449	5'	Potri.009G073900	
239-459	Chr01	6325550	6325949	5'	Potri.001G079900	transcription factor activity regulation of transcription, DNA-dependent
241-439	Chr05	21341000	21341349	5'	Potri.005G196800	
241-439	Chr05	21341000	21341349	5'	Potri.005G196900	
243-472	Chr11	1899950	1900399	5'	Potri.011G022800	
24-41	Chr01	30759500	30759949	5'	Potri.001G304600	diacylglycerol O-acyltransferase activity
248-264	Chr10	15666850	15667249	5'	Potri.010G147400	
251-186	Chr07	13020500	13020899	5'	Potri.007G106100	DNA binding
253-110	Chr14	6633450	6633749	5'	Potri.014G084300	
254-412	Chr10	19749750	19750149	5'	Potri.010G207300	nucleotide binding tRNA ligase activity ATP binding
256-515	Chr06	21507850	21508199	5'	Potri.006G200000	
257-279	Chr10	20360900	20361299	5'	Potri.010G216800	nucleic acid binding
258-308	Chr10	18081550	18081949	5'	Potri.010G183600	copper ion binding oxidation reduction oxidoreductase activity
265-377	Chr17	4116650	4116949	5'	Potri.017G049300	
270-207	Chr05	22006000	22006349	5'	Potri.005G205400	DNA binding regulation of transcription, DNA-dependent
27-47	Chr10	13644350	13644849	5'	Potri.010G119400	catalytic activity membrane catalytic activity membrane catalytic activity membrane catalytic activity trehalose biosynthesis catalytic activity trehalose biosynthesis
276-633	Chr12	14437950	14438599	5'	Potri.012G126100	

277-299	Chr17	14437500	14437949	5'	Potri.017G135200	
277-299	Chr17	14437500	14437949	5'	Potri.017G135300	
278-371	Chr04	4632500	4632949	5'	Potri.004G057800	
279-697	Chr03	13628750	13629199	5'	Potri.003G112700	oxidoreductase activity oxidation reduction copper ion binding
2807-502	Chr07	1953450	1953699	5'	Potri.007G026100	
28-31	Chr01	8442800	8443299	5'	Potri.001G106100	GTP binding GTP catabolism protein polymerization protein complex
283-266	Chr12	2921700	2922149	5'	Potri.012G032900	transcription factor activity regulation of transcription, DNA-dependent
288-420	Chr14	8584450	8584849	5'	Potri.014G109900	
292-389	Chr04	853200	853599	5'	Potri.004G013400	prephenate dehydratase activity L-phenylalanine biosynthesis
297-323	Chr14	8982050	8982399	5'	Potri.014G115200	calcium ion binding signal transduction extracellular matrix calcium ion binding signal transduction extracellular matrix
299-388	Chr08	8930900	8931299	5'	Potri.008G134300	
299-388	Chr08	8930900	8931299	5'	Potri.008G134500	
300-573	Chr03	16608050	16608499	5'	Potri.003G152200	
302-355	Chr10	22351250	22351699	5'	Potri.010G251700	RNA binding
302-355	Chr10	22351250	22351699	5'	Potri.010G251800	zinc ion binding intracellular zinc ion binding intracellular
3033-961	Chr10	4424900	4425149	5'	Potri.010G303400	
303-477	Chr15	2079100	2079599	5'	Potri.015G026600	
308-724	Chr06	11960300	11960699	5'	Potri.006G141700	cysteine-type peptidase activity proteolysis and peptidolysis
311-364	Chr10	15619250	15619699	5'	Potri.010G146700	catalytic activity
314-203	Chr10	10139000	10139399	5'	Potri.010G074800	cellulose synthase (UDP-forming) activity cellulose biosynthesis
315-784	Chr18	5462400	5462899	5'	Potri.018G052500	
316-173	Chr17	13241150	13241549	5'	Potri.017G117900	
317-674	Chr02	10835950	10836249	5'	Potri.002G145200	Golgi membrane Golgi cis-face ER to Golgi transport integral to membrane
318-457	Chr06	4141900	4142349	5'	Potri.006G056800	
320-137	Chr05	9511600	9511999	5'	Potri.005G122300	ATP binding
321-677	Chr06	2914000	2914399	5'	Potri.006G040800	catalytic activity metabolism catalytic activity metabolism catalytic activity
33-63	Chr14	1690800	1691299	5'	Potri.014G017500	
3382-883	Chr14	5776100	5776299	5'	Potri.014G071200	
338-569	Chr15	14063350	14063749	5'	Potri.015G129000	ATP binding nucleobase, nucleoside, nucleotide kinase activity
341-548	Chr02	10482700	10483249	5'	Potri.002G141200	transcription factor activity regulation of transcription, DNA-dependent
346-291	Chr09	9456450	9456799	5'	Potri.009G109900	catalytic activity metabolism catalytic activity metabolism
346-291	Chr09	9456450	9456799	5'	Potri.009G110000	
347-436	Chr05	4105550	4106049	5'	Potri.005G058300	similar to GT48; 1,3-beta-glucan synthase activity beta-1,3 glucan biosynthesis
348-666	Chr03	12617450	12617849	5'	Potri.003G100200	3-beta-hydroxy-delta5-steroid dehydrogenase activity
34-94	Chr08	9217450	9217999	5'	Potri.008G138300	integral to membrane
351-598	Chr11	17257400	17258099	5'	Potri.011G153800	pyrophosphatase activity cytoplasm pyrophosphatase activity
355-446	Chr07	791650	792099	5'	Potri.007G010100	
355-446	Chr07	791650	792099	5'	Potri.007G010200	drug transporter activity antiporter activity multidrug transport
361-652	Chr16	386550	386949	5'	Potri.016G007900	iron ion binding electron carrier activity
364-325	Chr02	1992350	1992749	5'	Potri.002G031000	transcription factor activity sequence-specific DNA binding
36-50	Chr19	884850	885199	5'	Potri.019G007500	
366-545	Chr04	7566700	7567049	5'	Potri.004G089100	protein kinase activity ATP binding
369-572	Chr13	2604400	2604749	5'	Potri.013G038100	
372-434	Chr06	24547700	24548199	5'	Potri.006G236200	regulation of transcription, DNA-dependent nucleus regulation of transcription, DNA-dependent nucleus
374-313	Chr03	14883000	14883349	5'	Potri.003G128700	oxidoreductase activity metabolism oxidoreductase activity metabolism oxidoreductase activity metabolism
375-132	Chr13	214050	214449	5'	Potri.013G003100	
376-332	Chr04	9452950	9453249	5'	Potri.004G107200	DNA binding regulation of transcription, DNA-dependent
377-613	Chr01	25589250	25589799	5'	Potri.001G245200	catalytic activity
379-671	Chr16	14184350	14184799	5'	Potri.016G139800	proton-transporting two-sector ATPase complex, catalytic domain
387-474	Chr17	9486150	9486599	5'	Potri.017G081100	ATP binding ATPase activity

388-98	Chr18	14305200	14305599	5'	Potri.018G118900	
388-98	Chr18	14305200	14305599	5'	Potri.018G119000	
390-648	Chr10	15661750	15662399	5'	Potri.010G147300	
3906-722	Chr18	213900	214099	5'	Potri.018G003100	ATP binding kinase activity phosphorylation ATP binding kinase activity phosphorylation ATP binding kinase activity phosphorylation ATP binding kinase activity phosphorylation
392-511	Chr14	11157500	11158049	5'	Potri.014G146200	
395-509	Chr10	17044450	17044849	5'	Potri.010G167500	DNA binding
39-80	scaffold_268	37450	37749	5'	Potri.T133400	protein binding
400-363	Chr10	6516650	6516999	5'	Potri.010G038300	transferase activity, transferring glycosyl groups
401-468	Chr14	7547450	7547849	5'	Potri.014G096300	DNA binding
402-327	Chr14	8650050	8650449	5'	Potri.014G110900	
403-316	Chr10	15186700	15187049	5'	Potri.010G139200	iron ion binding electron carrier activity oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
40-39	Chr01	26813200	26813749	5'	Potri.001G259100	oxidoreductase activity metabolism oxidation reduction aldenhyde dehydrogenase [NAD(P)] activity
409-632	Chr10	19200950	19201399	5'	Potri.010G199900	transferase activity, transferring glycosyl groups membrane
411-272	Chr05	25415250	25415699	5'	Potri.005G252000	protein kinase activity ATP binding
412-501	Chr15	14609550	14609949	5'	Potri.015G136400	transcription factor activity regulation of transcription, DNA-dependent
413-537	Chr15	8343450	8343849	5'	Potri.015G060900	
415-237	Chr12	10403150	10403649	5'	Potri.012G078300	
424-428	Chr05	3284700	3284999	5'	Potri.005G044600	
424-428	Chr05	3284700	3284999	5'	Potri.005G044700	
427-978	Chr06	24272450	24272849	5'	Potri.006G232800	
428-223	Chr14	1756300	1756699	5'	Potri.014G018900	serine-type endopeptidase activity proteolysis and peptidolysis
430-607	Chr19	11651750	11652049	5'	Potri.019G083600	DNA binding regulation of transcription, DNA-dependent
431-215	Chr09	4828200	4828549	5'	Potri.009G041000	
432-427	Chr08	13394250	13394699	5'	Potri.008G192400	ATP binding hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
432-427	Chr08	13394250	13394699	5'	Potri.008G192500	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor
43-99	Chr12	4913800	4914099	5'	Potri.012G051000	
4402-289	Chr17	12326700	12326899	5'	Potri.017G106000	
4402-289	Chr17	12326700	12326899	5'	Potri.017G106100	
443-467	Chr15	2344150	2344549	5'	Potri.015G029000	protein binding
448-482	Chr03	9420750	9421149	5'	Potri.003G066100	
451-441	Chr02	4075750	4076049	5'	Potri.002G060000	
454-391	Chr06	26559250	26559649	5'	Potri.006G262300	protein complex assembly cytochrome c oxidase biogenesis
455-473	Chr01	50018550	50019049	5'	Potri.001G466900	
45-67	Chr14	12502450	12503049	5'	Potri.014G159500	nucleic acid binding zinc ion binding nucleic acid binding zinc ion binding
457-238	Chr03	18552000	18552499	5'	Potri.003G177300	two-component signal transduction system (phosphorelay) two-component response regulator activity
457-238	Chr03	18552000	18552499	5'	Potri.003G177400	two-component response regulator activity two-component signal transduction system (phosphorelay)
460-651	Chr10	12616400	12617099	5'	Potri.010G106200	hydrolase activity
462-802	Chr08	11175400	11175899	5'	Potri.008G164200	
464-626	Chr01	47248600	47248999	5'	Potri.001G440900	protein binding nucleus DNA binding protein binding nucleus DNA binding protein binding nucleus DNA binding protein dimerization activity
467-220	Chr17	11722850	11723149	5'	Potri.017G099800	DNA binding
471-456	Chr01	13671650	13671949	5'	Potri.001G163500	RNA binding
474-913	Chr01	22504000	22504399	5'	Potri.001G219100	DNA binding
4758-981	Chr01	8142000	8142199	5'	Potri.001G102000	metallopeptidase activity zinc ion binding proteolysis and peptidolysis
477-542	Chr10	15511550	15511949	5'	Potri.010G144500	peptidyl-prolyl cis-trans isomerase activity protein folding peptidyl-prolyl cis-trans isomerase activity
48-123	Chr05	11184200	11184649	5'	Potri.005G136600	regulation of actin polymerization cytoskeleton regulation of actin polymerization cytoskeleton
485-825	Chr08	11666300	11666649	5'	Potri.008G170800	
4869-521	Chr18	14909700	14910049	5'	Potri.018G126700	protein binding
487-301	Chr05	744200	744499	5'	Potri.005G009700	
494-647	Chr17	5907650	5907999	5'	Potri.017G062900	peroxidase activity heme binding response to oxidative stress oxidation reduction
495-248	Chr13	7765350	7765999	5'	Potri.013G083600	

495-452	Chr13	7765350	7765999	5'	Potri.013G083600	peroxidase activity heme binding response to oxidative stress oxidation reduction
497-390	Chr01	35639700	35640199	5'	Potri.001G349900	protein kinase activity ATP binding protein amino acid phosphorylation protein binding
499-851	Chr01	33513400	33513799	5'	Potri.001G331500	protein binding
504-547	Chr04	451250	451599	5'	Potri.004G007200	
504-547	Chr04	451250	451599	5'	Potri.004G007300	
510-831	Chr08	6523850	6524249	5'	Potri.008G102700	
512-856	Chr17	13886150	13886499	5'	Potri.017G127000	protein binding ADP binding ATP binding apoptosis protein binding ADP binding
513-786	Chr19	5273300	5273699	5'	Potri.019G044500	
514-963	Chr07	4298300	4298949	5'	Potri.007G046700	microtubule motor activity ATP binding microtubule-based movement
518-255	Chr06	20550150	20550499	5'	Potri.006G190700	protein kinase activity ATP binding protein amino acid phosphorylation
520-454	Chr18	12864400	12864749	5'	Potri.018G100900	ATP binding
529-788	Chr17	10133000	10133699	5'	Potri.017G084200	structural constituent of ribosome intracellular ribosome protein biosynthesis
534-432	Chr06	9433400	9433799	5'	Potri.006G119300	serine carboxypeptidase activity proteolysis and peptidolysis serine carboxypeptidase activity
53-73	Chr08	1030100	1030499	5'	Potri.008G019400	[vesicle-mediated transport] integral to membrane vesicle-mediated transport
5376-442	Chr05	9629400	9629549	5'	Potri.005G123600	
537-860	Chr18	3264500	3264899	5'	Potri.018G039500	
541-277	Chr09	11894850	11895199	5'	Potri.009G151300	metalloendopeptidase activity proteolysis and peptidolysis metalloendopeptidase activity
543-739	Chr09	8953700	8954199	5'	Potri.009G101500	fucosyltransferase activity protein amino acid glycosylation membrane fucosyltransferase activity
549-892	Chr06	24063000	24063349	5'	Potri.006G229900	
550-479	Chr05	3196050	3196349	5'	Potri.005G043400	3-dehydroquinate dehydratase activity shikimate 5-dehydrogenase activity oxidation reduction
556-623	Chr10	339450	339799	5'	Potri.010G003800	
561-290	Chr15	2077650	2078099	5'	Potri.015G026600	
563-163	Chr06	8089450	8089799	5'	Potri.006G104600	
5645-811	Chr06	17954600	17954749	5'	Potri.006G171700	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor
566-965	Chr14	10523000	10523299	5'	Potri.014G137600	
567-350	Chr03	16863850	16864299	5'	Potri.003G156200	protein kinase activity ATP binding protein amino acid phosphorylation protein binding
57-12	Chr11	11522150	11522649	5'	Potri.011G094800	
573-614	Chr10	11248900	11249449	5'	Potri.010G088600	cellular morphogenesis ATP binding cellular morphogenesis ATP binding cellular morphogenesis ATP binding
573-614	Chr10	11248900	11249449	5'	Potri.010G088800	copper ion binding amine oxidase activity quinone binding amine metabolism
579-398	Chr05	23196750	23197049	5'	Potri.005G220300	
581-336	Chr07	1126700	1127049	5'	Potri.007G014800	microtubule motor activity ATP binding microtubule-based movement
584-785	Chr05	2018650	2019249	5'	Potri.005G027600	cellulose synthase (UDP-forming) activity cellulose biosynthesis membrane
586-719	Chr10	14263550	14263899	5'	Potri.010G126800	
590-959	Chr03	4034650	4034999	5'	Potri.003G032100	protein kinase activity ATP binding protein amino acid phosphorylation
591-701	Chr03	11595400	11595749	5'	Potri.003G088400	nucleoside metabolism
592-692	Chr06	10734900	10735249	5'	Potri.006G131000	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity
597-886	Chr01	30775400	30775699	5'	Potri.001G304800	O-methyltransferase activity
600-806	Chr17	9071100	9071499	5'	Potri.017G077300	protein kinase activity ATP binding protein amino acid phosphorylation
601-870	Chr07	1126050	1126549	5'	Potri.007G014800	microtubule motor activity ATP binding microtubule-based movement
603-522	Chr01	2089900	2090249	5'	Potri.001G028100	
604-997	Chr08	5283350	5283749	5'	Potri.008G084000	microtubule motor activity ATP binding microtubule-based movement
609-741	Chr04	18385550	18385899	5'	Potri.004G162800	
61-19	Chr16	2638700	2639149	5'	Potri.016G042600	protein binding
61-19	Chr16	2638700	2639149	5'	Potri.016G042700	
612-827	Chr10	18120800	18121249	5'	Potri.010G184000	
612-827	Chr10	18120800	18121249	5'	Potri.010G184100	nucleic acid binding ATP binding ATP dependent helicase activity
613-382	Chr14	2682150	2682499	5'	Potri.014G032800	hydrolase activity metabolism hydrolase activity metabolism hydrolase activity metabolism hydrolase activity
616-284	Chr16	4093700	4094149	5'	Potri.016G059300	activity metabolism
623-717	Chr14	2682700	2683149	5'	Potri.014G032800	hydrolase activity metabolism hydrolase activity metabolism hydrolase activity metabolism hydrolase activity
628-820	Chr10	20408650	20408999	5'	Potri.010G217700	activity metabolism

629-596	Chr04	21009800	21010149	5'	Potri.004G197500	protein kinase activity ATP binding protein amino acid phosphorylation
632-513	Chr19	14872850	14873149	5'	Potri.019G120800	hydrolase activity, acting on ester bonds lipid metabolism hydrolase activity, acting on ester bonds lipid metabolism
6388-814	scaffold_394	10150	10249	5'	Potri.T143600	hydrolase activity, acting on ester bonds lipid metabolism hydrolase activity, acting on ester bonds lipid metabolism
640-923	Chr01	7342150	7342499	5'	Potri.001G093300	biosynthesis cellular metabolism oxidoreductase activity metabolism
64-218	Chr12	15120000	15120449	5'	Potri.012G134500	
642-669	Chr02	3627150	3627499	5'	Potri.002G054400	
648-853	Chr14	11156650	11156999	5'	Potri.014G146200	
650-307	Chr15	13839400	13839799	5'	Potri.015G124800	membrane
653-871	Chr01	7634550	7634949	5'	Potri.001G097200	
655-715	Chr09	7137300	7137649	5'	Potri.009G073000	catalytic activity
664-813	Chr05	22367300	22367649	5'	Potri.005G209900	GTP binding intracellular GTP binding intracellular
665-580	Chr08	5798550	5799049	5'	Potri.008G092700	transferase activity, transferring glycosyl groups
673-534	Chr06	1805850	1806299	5'	Potri.006G026700	calcium ion binding signal transduction extracellular matrix
67-64	Chr12	13890650	13891099	5'	Potri.012G118500	protein binding
676-956	Chr16	3334000	3334449	5'	Potri.016G052000	
676-956	Chr16	3334000	3334449	5'	Potri.016G052100	
676-956	Chr16	3334000	3334449	5'	Potri.016G052200	
679-535	Chr19	1402300	1402699	5'	Potri.019G011800	protein kinase activity ATP binding protein amino acid phosphorylation
679-535	Chr19	1402300	1402699	5'	Potri.019G011900	protein amino acid phosphorylation protein kinase activity ATP binding
693-488	Chr15	11071050	11071349	5'	Potri.015G088100	ATP binding Holliday junction helicase activity DNA repair DNA recombination
698-602	Chr10	20208800	20209149	5'	Potri.010G214700	catalytic activity
700-675	Chr19	14797650	14797999	5'	Potri.019G120500	
700-675	Chr19	14797650	14797999	5'	Potri.019G120600	
70-129	scaffold_219	65050	65499	5'	Potri.T124900	
704-157	Chr01	48663100	48663499	5'	Potri.001G451900	oxidoreductase activity oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors
711-905	Chr19	11527100	11527499	5'	Potri.019G082000	integral to membrane
713-546	Chr02	16930700	16930999	5'	Potri.002G206100	
71-62	Chr06	2636000	2636499	5'	Potri.006G037900	
717-693	Chr11	16005350	16005749	5'	Potri.011G136400	GTP binding small GTPase mediated signal transduction intracellular
717-693	Chr11	16005350	16005749	5'	Potri.011G136400	
721-395	Chr01	6177300	6177649	5'	Potri.001G078100	
724-335	Chr01	28151450	28151799	5'	Potri.001G275100	
724-335	Chr01	28151450	28151799	5'	Potri.001G275200	protein binding phospholipid binding signal transduction intracellular
725-587	Chr13	16088450	16088849	5'	Potri.013G158500	transcription factor activity regulation of transcription, DNA-dependent
728-804	Chr10	15305650	15306149	5'	Potri.010G141400	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
730-686	Chr04	11276800	11277099	5'	Potri.004G120300	
730-686	Chr04	11276800	11277099	5'	Potri.004G120400	
731-404	Chr02	414150	414399	5'	Potri.002G007600	
734-893	Chr02	3336400	3336699	5'	Potri.002G050600	protein binding
749-740	Chr11	12709450	12709899	5'	Potri.011G104000	DNA binding protein dimerization activity DNA binding protein dimerization activity
749-740	Chr11	12709450	12709899	5'	Potri.011G104100	protein binding
755-839	Chr05	24668900	24669299	5'	Potri.005G240500	
755-839	Chr05	24668900	24669299	5'	Potri.005G240600	
769-527	Chr02	12433100	12433449	5'	Potri.002G164700	
770-383	Chr08	393500	394049	5'	Potri.008G007000	zinc ion binding intracellular zinc ion binding intracellular
770-383	Chr08	393500	394049	5'	Potri.008G007100	
773-346	Chr17	10821300	10821649	5'	Potri.017G090800	transferase activity, transferring glycosyl groups
776-636	Chr08	12326050	12326299	5'	Potri.008G180800	DNA binding
780-657	Chr10	20408000	20408349	5'	Potri.010G217700	
78-234	Chr08	13901950	13902349	5'	Potri.008G198700	catalytic activity

786-862	Chr16	14322300	14322649	5'	Potri.016G141700	transmembrane transport integral to membrane transmembrane transport
789-366	Chr19	14876450	14876799	5'	Potri.019G120900	similar to arabino galactan protein (AGP)
793-987	Chr09	9031850	9032299	5'	Potri.009G102800	protein binding
79-87	Chr15	11707000	11707549	5'	Potri.015G096800	
799-920	Chr11	16915100	16915399	5'	Potri.011G150700	structural constituent of ribosome ribosome protein biosynthesis
800-570	Chr13	6877450	6877749	5'	Potri.013G079100	
801-915	Chr10	9644750	9645199	5'	Potri.010G069200	
801-915	Chr10	9644750	9645199	5'	Potri.010G069300	structural molecule activity chloroplast
804-1000	Chr02	4987100	4987499	5'	Potri.002G072100	zinc ion binding oxidoreductase activity oxidation reduction zinc ion binding
806-431	Chr14	8505100	8505499	5'	Potri.014G108200	
80-93	Chr10	13352750	13353249	5'	Potri.010G115500	
809-872	Chr01	33279300	33279799	5'	Potri.001G328500	
812-393	Chr09	4424050	4424399	5'	Potri.009G034200	
813-979	Chr17	11894550	11894899	5'	Potri.017G102100	hydrolase activity, acting on ester bonds lipid metabolism hydrolase activity, acting on ester bonds
824-783	Chr18	3166150	3166499	5'	Potri.018G038100	transcription factor activity regulation of transcription, DNA-dependent
833-430	Chr10	8741350	8741599	5'	Potri.010G058100	threonine endopeptidase activity proteolysis during cellular protein catabolism
83-53	Chr02	3903200	3903749	5'	Potri.002G057800	intracellular protein transport cytoplasm intracellular protein transport cytoplasm
83-53	Chr02	3903200	3903749	5'	Potri.002G057900	integral to membrane
844-999	Chr15	13295600	13295949	5'	Potri.015G117900	metal ion transporter activity metal ion transport transmembrane transport
845-539	Chr06	27588650	27589049	5'	Potri.006G275900	DNA binding
84-61	Chr13	384000	384449	5'	Potri.013G005900	transferase activity, transferring glycosyl groups
850-948	Chr19	7264800	7265099	5'	Potri.019G048600	protein tyrosine/serine/threonine phosphatase activity protein amino acid dephosphorylation
867-500	Chr15	2061550	2061799	5'	Potri.015G026500	triacylglycerol lipase activity lipid metabolism triacylglycerol lipase activity lipid metabolism
87-119	Chr06	2357700	2358199	5'	Potri.006G034900	DNA binding
873-990	Chr04	1315550	1315949	5'	Potri.004G019800	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
884-738	Chr01	21849150	21849549	5'	Potri.001G215700	hydrolase activity protein binding hydrolase activity
896-653	Chr13	15238100	15238449	5'	Potri.013G147700	
909-876	Chr11	17007250	17007649	5'	Potri.011G151700	
911-711	Chr05	8526350	8526699	5'	Potri.005G110700	GTP binding small GTPase mediated signal transduction intracellular GTP binding
912-936	Chr04	4732000	4732399	5'	Potri.004G058900	
921-749	Chr10	12083450	12083749	5'	Potri.010G099100	
923-617	Chr15	9675650	9676149	5'	Potri.015G072500	
925-728	Chr17	11123750	11124149	5'	Potri.017G094300	
930-852	Chr13	181100	181399	5'	Potri.013G002400	metabolism
930-852	Chr13	181100	181399	5'	Potri.013G002500	
93-102	Chr03	21438300	21438649	5'	Potri.003G218300	
931-679	Chr12	561250	561549	5'	Potri.012G006400	O-methyltransferase activity methyltransferase activity protein dimerization activity
931-679	Chr12	561250	561549	5'	Potri.012G006500	
934-944	Chr03	13541650	13542099	5'	Potri.003G111800	
955-800	Chr16	12496150	12496599	5'	Potri.016G120100	magnesium ion binding thiamin pyrophosphate binding catalytic activity
971-628	Chr04	3257950	3258299	5'	Potri.004G042900	
986-940	Chr09	7367000	7367399	5'	Potri.009G076000	
991-507	Chr02	2315750	2316049	5'	Potri.002G036000	protein kinase activity ATP binding protein amino acid phosphorylation
993-489	scaffold_67	75050	75399	5'	Potri.T064300	
997-823	Chr06	21903800	21904199	5'	Potri.006G203600	
3-2	Chr01	42928500	42928999	3'	Potri.001G406800	
11-4	Chr19	12806550	12806999	3'	Potri.019G097700	protein binding
4-8	scaffold_93	123750	124449	3'	Potri.T082700	
9-14	Chr06	2399650	2400149	3'	Potri.006G035400	protein polymerization protein complex GTPase activity GTP binding
12-15	Chr02	10465000	10465549	3'	Potri.002G140900	DNA binding
1001-574	Chr02	12003350	12003699	3'	Potri.002G159700	Rho guanyl-nucleotide exchange factor activity

1003-898	Chr19	12034250	12034599	3'	Potri.019G088600	oxidoreductase activity oxidation reduction copper ion binding
1013-620	Chr17	13742700	13742999	3'	Potri.017G125500	protein binding
1014-762	scaffold_49	9800	10149	3'	Potri.T053300	
1023-986	Chr11	17450200	17450499	3'	Potri.011G156800	
1023-986	Chr11	17450200	17450499	3'	Potri.011G156800	
104-128	Chr10	17792600	17793099	3'	Potri.010G179000	nucleic acid binding
1047-618	Chr03	17549900	17550249	3'	Potri.003G164900	
1049-955	Chr05	8594100	8594449	3'	Potri.005G111900	
105-135	Chr02	7322700	7323149	3'	Potri.002G101400	
1051-937	Chr07	9249350	9249749	3'	Potri.007G070500	nucleotide binding tRNA ligase activity ATP binding amino acid activation
106-113	Chr13	15546200	15546799	3'	Potri.013G151400	integral to membrane
1063-903	Chr03	12025000	12025299	3'	Potri.003G093000	heat shock protein binding
109-144	Chr19	11656850	11657349	3'	Potri.019G083700	transferase activity, transferring acyl groups metabolism acyltransferase activity
1093-600	Chr05	1876050	1876299	3'	Potri.005G025400	oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor oxidation reduction oxidation reduction
1096-586	Chr05	19974600	19974899	3'	Potri.005G183000	regulation of transcription, DNA-dependent nucleus regulation of transcription, DNA-dependent
1100-924	Chr01	17124450	17124749	3'	Potri.001G190300	RNA binding pseudouridine synthase activity pseudouridine synthesis RNA modification
111-217	Chr09	8644550	8645049	3'	Potri.009G095700	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
1120-504	Chr02	16422900	16423149	3'	Potri.002G202400	glycosyl compounds carbohydrate metabolism
1120-504	Chr02	16422900	16423149	3'	Potri.002G202300	biosynthesis sucrose metabolism biosynthesis sucrose metabolism biosynthesis
113-30	Chr01	48387200	48387749	3'	Potri.001G449600	intracellular protein transport intracellular protein transport intracellular protein binding
1146-770	Chr04	16042950	16043199	3'	Potri.004G138600	
1167-764	Chr01	28395450	28395749	3'	Potri.001G277400	
1174-769	Chr10	7379300	7379649	3'	Potri.010G042900	
1177-731	Chr07	2025550	2025749	3'	Potri.007G026700	DNA binding ATP binding nucleic acid binding helicase activity DNA binding
1182-519	Chr13	13106900	13107199	3'	Potri.013G117900	endocytosis membrane endocytosis membrane
1216-939	Chr08	13188100	13188399	3'	Potri.008G190000	
1222-743	Chr04	148100	148699	3'	Potri.004G001900	protein kinase activity ATP binding protein amino acid phosphorylation
122-306	Chr10	13174400	13174849	3'	Potri.010G113400	heat shock protein binding
123-148	Chr07	14915350	14915799	3'	Potri.007G136900	DNA binding
1243-761	Chr11	15236800	15237099	3'	Potri.011G127400	pyridoxal phosphate binding amino acid metabolism pyridoxal phosphate binding
1243-761	Chr11	15236800	15237099	3'	Potri.011G127300	ADP binding ATP binding apoptosis ADP binding ATP binding apoptosis
1260-563	Chr12	3994300	3994649	3'	Potri.012G044100	phosphatidylinositol phosphate kinase activity phosphatidylinositol metabolism
1261-249	Chr02	6244350	6244599	3'	Potri.002G088100	
126-131	Chr01	30639700	30640149	3'	Potri.001G302700	
126-131	Chr01	30639700	30640149	3'	Potri.001G302800	
1285-630	Chr01	7903900	7904149	3'	Potri.001G099800	DNA binding
1294-988	Chr12	390600	390949	3'	Potri.012G004600	mitochondrial inner membrane protein transporter activity protein transport
1297-672	Chr13	1160500	1160799	3'	Potri.013G017700	metal ion binding metal ion transport metal ion binding metal ion transport metal ion binding metal ion transport
1319-799	Chr06	11491150	11491499	3'	Potri.006G138000	transferase activity, transferring groups other than amino-acyl groups metabolism
1327-904	Chr03	14428350	14428649	3'	Potri.003G122000	enzyme inhibitor activity pectinesterase activity enzyme inhibitor activity pectinesterase activity
133-271	Chr13	14495950	14496399	3'	Potri.013G134200	integral to membrane
134-190	Chr03	18524000	18524599	3'	Potri.003G176900	biosynthesis transferase activity, transferring groups other than amino-acyl groups
135-196	Chr06	25753000	25753499	3'	Potri.006G251800	transcription factor activity sequence-specific DNA binding protein dimerization activity
1354-610	Chr14	1366050	1366299	3'	Potri.014G013300	microtubule motor activity ATP binding microtubule-based movement microtubule motor activity
1354-610	Chr14	1366050	1366299	3'	Potri.014G013100	protein binding
1359-992	Chr14	10402650	10402999	3'	Potri.014G136100	protein binding
1359-992	Chr14	10402650	10402999	3'	Potri.014G136000	
136-140	Chr05	19000400	19000849	3'	Potri.005G174500	

137-76	Chr05	23591950	23592399	3'	Potri.005G225700	sugar binding carbohydrate metabolism sugar binding carbohydrate metabolism
1395-709	Chr12	15405950	15406399	3'	Potri.012G139900	GTP binding small GTPase mediated signal transduction nucleus cytoplasm
140-104	Chr04	3932100	3932549	3'	Potri.004G050800	
140-104	Chr04	3932100	3932549	3'	Potri.004G050700	
142-146	Chr02	9491350	9491749	3'	Potri.002G126200	sterol binding
1433-918	Chr14	8138850	8139249	3'	Potri.014G103500	protein kinase activity ATP binding protein amino acid phosphorylation
1433-918	Chr14	8138850	8139249	3'	Potri.014G103400	
145-213	Chr09	6278400	6278699	3'	Potri.009G060700	
148-169	Chr06	19365900	19366349	3'	Potri.006G179400	protein kinase activity ATP binding protein amino acid phosphorylation
149-139	Chr19	14280450	14280799	3'	Potri.019G114900	
1498-634	Chr10	12722200	12722549	3'	Potri.010G107400	intracellular zinc ion binding intracellular zinc ion binding
1506-281	Chr07	11318600	11318849	3'	Potri.007G087100	integral to mitochondrial inner membrane
15-16	Chr02	1201750	1202199	3'	Potri.002G020300	
1537-843	Chr17	14771400	14771799	3'	Potri.017G139700	autophagy cytoplasm autophagy cytoplasm
161-156	Chr04	7168200	7168799	3'	Potri.004G086200	GTP binding small GTPase mediated signal transduction signal transducer activity
161-156	Chr04	7168200	7168799	3'	Potri.004G086100	protein kinase activity ATP binding protein amino acid phosphorylation protein binding
1613-884	Chr16	4717050	4717299	3'	Potri.016G066500	
162-83	Chr10	11510100	11510599	3'	Potri.010G091600	mitosis
164-172	Chr01	4189450	4189899	3'	Potri.001G055200	integral to membrane transmembrane transporter activity transmembrane transport
165-145	Chr03	13294700	13295099	3'	Potri.003G109300	microtubule motor activity ATP binding microtubule-based movement
1708-824	Chr02	22904500	22904949	3'	Potri.002G235500	peroxidase activity heme binding response to oxidative stress oxidation reduction peroxidase activity heme binding response to oxidative stress oxidation reduction
1710-84	Chr05	21241600	21241899	3'	Potri.005G195600	electron carrier activity iron-sulfur cluster binding
1737-676	Chr09	12014600	12014849	3'	Potri.009G153800	ubiquitin thioesterase activity
175-175	Chr01	48338550	48338999	3'	Potri.001G449000	
179-239	Chr07	14985900	14986249	3'	Potri.007G138100	transcription factor activity regulation of transcription, DNA-dependent
181-187	Chr04	3686400	3686899	3'	Potri.004G047900	
187-194	Chr17	14219350	14219649	3'	Potri.017G131800	aspartic-type endopeptidase activity proteolysis and peptidolysis
189-727	Chr03	18003850	18004349	3'	Potri.003G169900	transcription factor activity regulation of transcription, DNA-dependent
196-108	Chr05	24045950	24046349	3'	Potri.005G232100	acid-D-amino acid ligase activity
198-252	Chr14	4215900	4216349	3'	Potri.014G053600	
200-149	Chr02	3636900	3637449	3'	Potri.002G054500	
2007-403	Chr04	1580100	1580349	3'	Potri.004G023800	protein kinase activity ATP binding protein amino acid phosphorylation
208-125	Chr07	12617400	12617949	3'	Potri.007G100400	
208-125	Chr07	12617400	12617949	3'	Potri.007G100500	
209-292	Chr10	13953300	13953949	3'	Potri.010G122900	nucleotide binding ATP binding ligase activity, forming aminoacyl-tRNA and related compounds
211-426	Chr09	6307700	6308149	3'	Potri.009G061100	drug transporter activity antiporter activity multidrug transport transmembrane transport
211-426	Chr09	6307700	6308149	3'	Potri.009G061000	protein binding
212-236	Chr04	9283750	9284149	3'	Potri.004G105600	
212-236	Chr04	9283750	9284149	3'	Potri.004G105400	
2123-975	Chr03	11139750	11140099	3'	Potri.003G083800	
215-293	Chr18	15245250	15245699	3'	Potri.018G131200	catalytic activity metabolism catalytic activity metabolism catalytic activity metabolism
217-265	Chr15	603850	604249	3'	Potri.015G09200	
2218-907	Chr04	8172450	8172699	3'	Potri.004G095100	zinc ion binding nucleic acid binding zinc ion binding nucleic acid binding zinc ion binding nucleic acid binding
22-20	Chr03	15937400	15938149	3'	Potri.003G142700	proteasome regulatory particle (sensu Eukarya) proteolysis and peptidolysis
225-379	Chr10	16724900	16725299	3'	Potri.010G162100	transferase activity, transferring glycosyl groups
2241-339	Chr01	49641200	49641399	3'	Potri.001G462200	UDP-N-acetylmuramate dehydrogenase activity oxidoreductase activity FAD binding
226-517	Chr08	10829800	10830249	3'	Potri.008G159600	
232-319	Chr01	25778800	25779299	3'	Potri.001G247500	magnesium ion binding phosphopyruvate hydratase activity glycolysis
234-369	Chr09	7220000	7220449	3'	Potri.009G074000	
234-369	Chr09	7220000	7220449	3'	Potri.009G073800	[6-phospho]fructo-2-kinase activity ATP binding fructose metabolism
2363-718	Chr16	3665850	3666099	3'	Potri.016G055800	

238-650	Chr08	8368650	8368999	3'	Potri.008G128500	DNA binding
240-438	Chr10	7811000	7811399	3'	Potri.010G046700	
244-51	Chr06	11824250	11824649	3'	Potri.006G140800	
248-264	Chr10	15666850	15667249	3'	Potri.010G147300	
2527-844	Chr09	9339800	9340099	3'	Potri.009G107700	oxidoreductase activity
2527-844	Chr09	9339800	9340099	3'	Potri.009G107800	
252-837	Chr07	638550	638899	3'	Potri.007G008200	DNA binding nucleus regulation of transcription, DNA-dependent transcription factor activity
253-110	Chr14	6633450	6633749	3'	Potri.014G084200	Rho guanyl-nucleotide exchange factor activity
256-515	Chr06	21507850	21508199	3'	Potri.006G199900	
258-308	Chr10	18081550	18081949	3'	Potri.010G183500	oxidoreductase activity
260-159	Chr12	3272750	3273049	3'	Potri.012G036700	
261-106	Chr19	12702400	12702749	3'	Potri.019G095900	intracellular protein binding
263-737	Chr01	30577800	30578249	3'	Potri.001G301900	protein binding
266-349	Chr06	27636450	27636949	3'	Potri.006G276800	
267-411	Chr14	7094200	7094749	3'	Potri.014G090100	exocytosis vesicle docking cytoplasm exocytosis vesicle docking cytoplasm
279-697	Chr03	13628750	13629199	3'	Potri.003G112600	beta-fructofuranosidase activity sucrose alpha-glucosidase activity
2834-579	Chr15	14283500	14283749	3'	Potri.015G132500	2 iron, 2 sulfur cluster binding intracellular membrane-bound organelle
287-861	Chr09	3605350	3605699	3'	Potri.009G024600	
296-664	Chr03	20082600	20083199	3'	Potri.003G198100	electron carrier activity protein disulfide oxidoreductase activity
297-323	Chr14	8982050	8982399	3'	Potri.014G115300	oxidoreductase activity metabolism
299-388	Chr08	8930900	8931299	3'	Potri.008G134400	acid-D-amino acid ligase activity
300-573	Chr03	16608050	16608499	3'	Potri.003G152300	
304-326	Chr08	4091900	4092249	3'	Potri.008G067400	protein binding DNA binding protein binding DNA binding protein binding
305-311	Chr01	12803350	12803799	3'	Potri.001G155100	transcription factor activity sequence-specific DNA binding regulation of transcription
309-435	Chr10	5687250	5687699	3'	Potri.010G033500	voltage-dependent ion-selective channel activity anion transport
31-35	Chr07	8976550	8977199	3'	Potri.007G068600	
31-35	Chr07	8976550	8977199	3'	Potri.007G068700	
314-203	Chr10	10139000	10139399	3'	Potri.010G074900	protein binding
3156-461	Chr13	3199250	3199449	3'	Potri.013G045200	transcription factor activity regulation of transcription, DNA-dependent
319-324	Chr06	23159500	23159899	3'	Potri.006G219200	copper ion transporter activity copper ion transmembrane transport
320-137	Chr05	9511600	9511999	3'	Potri.005G122200	DNA binding
322-229	Chr08	8867700	8868149	3'	Potri.008G133500	
326-506	Chr10	1754350	1754699	3'	Potri.010G014100	
327-583	Chr12	12459050	12459499	3'	Potri.012G099100	
332-753	Chr08	4599150	4599649	3'	Potri.008G073900	protein kinase activity ATP binding protein amino acid phosphorylation
333-465	Chr05	15347850	15348249	3'	Potri.005G155600	cysteine-type peptidase activity proteolysis and peptidolysis
334-367	Chr18	2164350	2164749	3'	Potri.018G027200	
334-367	Chr18	2164350	2164749	3'	Potri.018G027100	
339-443	Chr14	6309500	6310249	3'	Potri.014G078500	
342-405	Chr10	13084100	13084399	3'	Potri.010G112300	protein kinase activity ATP binding protein amino acid phosphorylation
344-599	Chr08	1479600	1479949	3'	Potri.008G027500	metalloendopeptidase activity zinc ion binding proteolysis and peptidolysis
35-34	scaffold_20	748250	748799	3'	Potri.T004700	serine-type endopeptidase activity proteolysis and peptidolysis protein self binding
35-34	scaffold_20	748250	748799	3'	Potri.T004600	triacylglycerol lipase activity lipid metabolism triacylglycerol lipase activity lipid metabolism
355-446	Chr07	791650	792099	3'	Potri.007G010000	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
360-221	Chr12	12206150	12206549	3'	Potri.012G095500	biosynthesis
361-652	Chr16	386550	386949	3'	Potri.016G008000	
368-269	Chr07	1036250	1036549	3'	Potri.007G013600	
368-269	Chr07	1036250	1036549	3'	Potri.007G013400	isomerase activity
368-269	Chr07	1036250	1036549	3'	Potri.007G013500	DNA binding transcription initiation nucleus nucleosome nucleosome assembly
373-499	Chr02	23698600	23699049	3'	Potri.002G245500	
374-313	Chr03	14883000	14883349	3'	Potri.003G128600	transporter activity transport membrane transporter activity transport membrane

379-671	Chr16	14184350	14184799	3'	Potri.016G139900	protein binding signal transducer activity response to light protein binding
38-96	Chr10	18071750	18072249	3'	Potri.010G183400	protein kinase activity ATP binding protein amino acid phosphorylation
391-763	Chr06	3617100	3617449	3'	Potri.006G050200	actin binding cytoskeleton organization and biogenesis actin binding
391-763	Chr06	3617100	3617449	3'	Potri.006G050100	protein binding
397-385	Chr04	1430850	1431399	3'	Potri.004G021600	iron ion binding electron carrier activity oxidoreductase activity, acting on paired donors,
403-316	Chr10	15186700	15187049	3'	Potri.010G139300	catalytic activity
405-561	Chr14	2622000	2622349	3'	Potri.014G031400	zinc ion binding
405-561	Chr14	2622000	2622349	3'	Potri.014G031500	RNA binding ribonuclease activity RNA binding ribonuclease activity
407-312	Chr06	22533500	22533799	3'	Potri.006G210800	nucleus
41-25	Chr12	10265200	10265649	3'	Potri.012G077300	structural constituent of ribosome protein biosynthesis ribosome
41-25	Chr12	10265200	10265649	3'	Potri.012G077200	DNA binding protein dimerization activity transcription factor activity
421-409	Chr02	6077950	6078399	3'	Potri.002G086400	recognition or rejection of self pollen protein amino acid phosphorylation
421-409	Chr02	6077950	6078399	3'	Potri.002G086300	acid-D-amino acid ligase activity
454-391	Chr06	26559250	26559649	3'	Potri.006G262400	ubiquitin thiolesterase activity ubiquitin-dependent protein catabolism
466-926	Chr08	431850	432249	3'	Potri.008G080000	protein binding
467-220	Chr17	11722850	11723149	3'	Potri.017G099700	catalytic activity metabolism catalytic activity metabolism catalytic activity metabolism
46-78	Chr02	1387550	1388049	3'	Potri.002G023100	photosynthesis photosystem I photosystem I reaction center photosynthesis
469-455	Chr13	10358000	10358399	3'	Potri.013G095800	protein kinase activity ATP binding protein amino acid phosphorylation
471-456	Chr01	13671650	13671949	3'	Potri.001G163400	thiolester hydrolase activity fatty acid biosynthesis thiolester hydrolase activity fatty acid biosynthesis
4758-981	Chr01	8142000	8142199	3'	Potri.001G101900	ATP binding kinase activity phosphorylation ATP binding kinase activity phosphorylation
478-425	Chr14	9444600	9445099	3'	Potri.014G122100	DNA-3-methyladenine glycosylase I activity base-excision repair DNA-3-methyladenine glycosylase I activity base-excision repair
484-663	Chr02	7610800	7611149	3'	Potri.002G104800	phosphotransferase activity, alcohol group as acceptor
487-301	Chr05	7444200	7444499	3'	Potri.005G009800	phosphotransferase activity, alcohol group as acceptor
500-471	Chr08	10287650	10288199	3'	Potri.008G151700	protein binding protein transporter activity intracellular protein transport
500-471	Chr08	10287650	10288199	3'	Potri.008G151600	methyltransferase activity metabolism methyltransferase activity metabolism
502-508	Chr04	19844000	19844399	3'	Potri.004G181800	transmembrane transporter activity transmembrane transport integral to membrane
504-547	Chr04	451250	451599	3'	Potri.004G007100	sugar binding recognition or rejection of self pollen protein amino acid phosphorylation
506-767	Chr05	22616850	22617549	3'	Potri.005G212800	phosphotransferase activity, alcohol group as acceptor
508-440	scaffold_93	42600	42999	3'	Potri.1082100	phosphotransferase activity, alcohol group as acceptor
510-831	Chr08	6523850	6524249	3'	Potri.008G102800	protein binding ADP binding ATP binding apoptosis protein binding ADP binding ATP binding
511-584	Chr08	7172000	7172799	3'	Potri.008G112300	zinc ion binding
511-584	Chr08	7172000	7172799	3'	Potri.008G112200	oxidoreductase activity oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors
517-170	Chr07	15039800	15040299	3'	Potri.007G139300	receptor activity endoplasmic reticulum lumen
523-261	Chr02	5225200	5225599	3'	Potri.002G075700	copper ion binding amine oxidase activity quinone binding amine metabolism oxidation reduction
523-261	Chr02	5225200	5225599	3'	Potri.002G075600	
526-566	Chr08	5388000	5388349	3'	Potri.008G085800	
5302-706	Chr11	14947750	14947849	3'	Potri.011G124600	
536-597	Chr08	10212000	10212399	3'	Potri.008G150900	
537-860	Chr18	3264500	3264899	3'	Potri.018G039400	
540-283	scaffold_29	334000	334249	3'	Potri.1023000	
543-739	Chr09	8953700	8954199	3'	Potri.009G101400	
548-424	Chr06	26238100	26238449	3'	Potri.006G258700	
54-85	Chr07	13259650	13260099	3'	Potri.007G109000	
55-58	Chr11	15531700	15532349	3'	Potri.011G129800	
556-623	Chr10	339450	339799	3'	Potri.010G003700	
563-163	Chr06	8089450	8089799	3'	Potri.006G104700	
566-965	Chr14	10523000	10523299	3'	Potri.014G137700	
566-965	Chr14	10523000	10523299	3'	Potri.014G137500	
573-614	Chr10	11248900	11249449	3'	Potri.010G088700	

766-765	Chr02	11686800	11687099	3'	Potri.002G155100	protein kinase activity ATP binding protein amino acid phosphorylation protein kinase activity ATP binding protein amino acid phosphorylation
766-765	Chr02	11686800	11687099	3'	Potri.002G155000	
775-619	Chr01	20338250	20338899	3'	Potri.001G204700	
787-801	Chr08	4479100	4479449	3'	Potri.008G072500	RNA binding translation initiation factor activity eukaryotic translation initiation factor 2 complex
787-801	Chr08	4479100	4479449	3'	Potri.008G072400	
79-87	Chr15	11707000	11707549	3'	Potri.015G096700	
800-570	Chr13	6877450	6877749	3'	Potri.013G079000	
802-747	Chr07	13517400	13517699	3'	Potri.007G113000	
803-629	Chr09	4447350	4447649	3'	Potri.009G034600	metal ion transporter activity metal ion transport transmembrane transport membrane
804-1000	Chr02	4987100	4987499	3'	Potri.002G072200	transmembrane transport integral to membrane
80-93	Chr10	13352750	13353249	3'	Potri.010G115400	
812-393	Chr09	4424050	4424399	3'	Potri.009G034100	
817-503	Chr09	4385950	4386249	3'	Potri.009G033800	
82-32	Chr10	7387500	7387899	3'	Potri.010G043000	
826-917	Chr11	561150	561449	3'	Potri.011G007600	aspartic-type endopeptidase activity proteolysis and peptidolysis lipid metabolism
827-841	Chr07	10078450	10078749	3'	Potri.007G076900	protein binding
832-958	Chr11	4437700	4438099	3'	Potri.011G051800	
832-958	Chr11	4437700	4438099	3'	Potri.011G051700	
842-712	Chr01	44558500	44558899	3'	Potri.001G420300	catalytic activity pyridoxal phosphate binding metabolism catalytic activity
848-874	Chr09	10744500	10744849	3'	Potri.009G132600	hydrolase activity, hydrolyzing O-glycosyl compounds carbohydrate metabolism
849-378	Chr08	9843550	9843849	3'	Potri.008G146200	
850-948	Chr19	7264800	7265099	3'	Potri.019G048500	
85-107	Chr07	2414000	2414449	3'	Potri.007G031800	
851-957	Chr05	4537500	4538049	3'	Potri.005G063100	
854-858	Chr07	1187800	1188099	3'	Potri.007G016000	protein binding
87-119	Chr06	2357700	2358199	3'	Potri.006G035000	ATP binding
873-990	Chr04	1315550	1315949	3'	Potri.004G019900	membrane transmembrane transport membrane transmembrane transport
879-658	Chr09	5198650	5198899	3'	Potri.009G046500	peptidyl-prolyl cis-trans isomerase activity protein folding
879-658	Chr09	5198650	5198899	3'	Potri.009G046400	
892-704	Chr17	15209750	15209999	3'	Potri.017G144400	ADP binding ATP binding apoptosis ADP binding ATP binding apoptosis
897-505	Chr08	3756300	3756649	3'	Potri.008G062200	
901-721	Chr07	14765500	14765799	3'	Potri.007G134500	DNA binding
90-216	Chr10	21452750	21453199	3'	Potri.010G233900	
922-976	Chr02	2581750	2582149	3'	Potri.002G040400	oxidoreductase activity oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors
922-976	Chr02	2581750	2582149	3'	Potri.002G040500	
928-748	Chr01	13110850	13111249	3'	Potri.001G157700	
930-852	Chr13	181100	181399	3'	Potri.013G002300	
936-938	Chr08	1409600	1409899	3'	Potri.008G026500	
939-406	Chr01	48253900	48254499	3'	Potri.001G448400	DNA binding regulation of transcription, DNA-dependent
94-109	Chr02	10341800	10342249	3'	Potri.002G139200	transferase activity, transferring glycosyl groups GPI anchor biosynthesis
94-109	Chr02	10341800	10342249	3'	Potri.002G139300	
949-288	Chr02	23759900	23760199	3'	Potri.002G246300	
96-55	Chr08	7397350	7397799	3'	Potri.008G115500	
97-185	Chr01	14581050	14581499	3'	Potri.001G172000	
973-846	Chr13	6683700	6683899	3'	Potri.013G077500	protein binding ATP binding hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
98-233	scaffold_1833	4350	4849	3'	Potri.T178700	metabolism transferase activity, transferring hexosyl groups
991-507	Chr02	2315750	2316049	3'	Potri.002G036100	
99-68	Chr01	9364650	9365099	3'	Potri.001G116700	
997-823	Chr06	21903800	21904199	3'	Potri.006G203500	protein kinase activity ATP binding

Wassim LAKHAL

Etude fonctionnelle de trois facteurs de transcription impliqués dans la formation de la paroi secondaire chez le peuplier

Les facteurs de transcription (FT) de la famille R2R3-MYB chez les plantes jouent un rôle important dans la formation de la paroi secondaire des cellules de bois, que ce soit en activant ou en réprimant leurs gènes cibles au sein d'un réseau régulationnel complexe. Dans ce travail, nous avons utilisé la transgénèse et l'immunoprécipitation de chromatine associée à un séquençage haut-débit (ChIP-SEQ) pour déterminer la fonction de 3 FT R2R3-MYB chez le peuplier. Les peupliers surexprimant MYB090 ont des rayons moins lignifiés ; les tiges présentent une réduction de croissance et de teneurs en lignines. MYB090 régule ses cibles à l'aide d'un motif très conservé, similaire au motif Gamyb. Ses cibles sont impliquées notamment dans la biosynthèse des lignines, cellulose et xylanes, constituants principaux des parois. Les plantes surexprimant MYB221-SRDX et MYB156 présentent une nette réduction de la lignification des parois de leurs fibres, associée à une réduction de croissance. MYB221 semble avoir pour cibles des gènes codant pour des enzymes du métabolisme, ainsi qu'un autre FT de type R2R3-MYB, dont la régulation passe par un motif conservé de type SMRE (Secondary wall MYB-Responsive Element). En conclusion, la combinaison des approches ChIP-SEQ et de transgénèse montre que MYB090 semble être un répresseur transcriptionnel de la lignification, notamment dans les rayons, et de la formation de la paroi secondaire. De même, MYB156 et MYB221 seraient également des répresseurs de la lignification, dans les fibres et les rayons. Cette thèse ouvre des perspectives sur l'établissement de réseaux de régulation transcriptionnelle de la formation de la paroi secondaire.

Mots clés : Facteur de transcription, lignines, paroi secondaire, cellulose, xylanes, ChIP-SEQ, peuplier

Functional study of three transcription factors involved in secondary cell wall formation in poplar

Plant R2R3-MYB transcription factors (TF) play an important role in secondary cell wall formation in wood cells, by activating or repressing their target genes within a complex regulatory network. Here, we used genetic engineering and chromatin immunoprecipitation technique, associated to next-generation sequencing (ChIP-SEQ) to determine the function of 3 R2R3-MYB TF in poplar. Plants overexpressing MYB090 had less lignified parenchyma rays. The stem growth and total lignin content were reduced. MYB090 regulates target genes through a highly conserved motif, similar to Gamyb. Its target genes are involved in lignin, cellulose and xylan biosynthesis, which are the major components of secondary cell wall. Poplars overexpressing MYB221-SRDX and MYB156 showed a decrease in fiber cell wall lignification, and a reduced growth. MYB221 have targets encoding for metabolic enzymes but also for another R2R3-MYB TF. MYB221 recognizes its target genes, most probably through SMRE (Secondary wall MYB-Responsive Element) conserved motif. In conclusion, the combination of ChIP-SEQ and genetic engineering approaches shows that MYB090 seems to be a transcriptional repressor of lignification, especially in parenchyma rays. MYB156 and MYB221 are also negative regulators of secondary cell wall lignification, in fibers and parenchyma rays. This work opens new avenues on the building of transcriptional regulatory networks involved in secondary cell wall formation.

Keywords : Transcription factor, lignins, secondary cell wall, cellulose, xylanes, ChIP-SEQ, poplar



INRA UR 0588
Unité Amélioration, Génétique et
Physiologie Forestières

