

Analyse de peupliers transgéniques sur-exprimant miR397a et caractérisation des sites d'insertion des transgènes

Université d'Orléans

Collegium sciences et techniques

IPT sciences biologique et chimie du vivant

MASTER SCIENCES BIOLOGIES 2^{ème} année

Spécialité : **BBMB**

Année 2012-2013

Audrey Chartrin

Organisme d'accueil

INRA – Institut National de Recherche Agronomique

Équipe d'accueil

Amélioration Génétique et Physiologie Forestières

Responsable

Jean-Charles Leplé - Chargé de Recherche

Co-encadrant

& Delphine Gourcilleau – Post doctorante

REMERCIEMENTS

Je tiens à remercier Gilles Pilate, directeur de l'unité Amélioration, Génétique et Physiologie Forestière, pour m'avoir acceptée dans son équipe.

Je remercie également Jean-Charles Leplé ainsi que Delphine Gourcilleau pour leur soutien, leur aide précieuse et leur patience tout du long de mon stage.

Merci de même à Marie-Claude Lesage-Descauses, Vanina Guerin, Nadège Millet et Nicolas Richet pour leur gentillesse et leurs précieux conseils.

Merci enfin à toutes les personnes que je n'ai pas nommées mais qui ont rendu cette expérience enrichissante et pleine d'intérêt.

SOMMAIRE

ABRÉVIATIONS.....	1
PRESENTATION SUCCINCTE DE L'ORGANISME D'ACCUEIL.....	2
INTRODUCTION	2
1) INTRODUCTION GENERALE	2
2) ETUDE BIBLIOGRAPHIQUE	3
a. <i>Les sites d'insertion et leur emplacement.....</i>	<i>3</i>
b. <i>Micro-ARN, laccases et lignification</i>	<i>5</i>
3) PRESENTATION DU THEME DU STAGE	6
a. <i>La mise au point d'une méthode permettant de déterminer les sites d'insertion de transgènes.....</i>	<i>6</i>
b. <i>L'étude d'expression de peupliers transgéniques sur-exprimant le micro ARN miR397a, et l'impact sur l'expression de gènes de laccases.....</i>	<i>7</i>
MATERIELS ET METHODES	8
1) CONSTRUCTIONS UTILISEES	8
2) MATERIEL VEGETAL.....	8
3) GENERALITES SUR LA CARACTERISATION DES SITES D'INSERTION	8
4) AMPLIFICATION DES SITES D'INSERTION	10
a. <i>Extractions d'ADN génomique.....</i>	<i>10</i>
b. <i>Digestion de l'ADN génomique et ligation des adaptateurs.....</i>	<i>10</i>
c. <i>PCR et migrations sur gel.....</i>	<i>10</i>
5) CLONAGE ET SEQUENÇAGE DES AMPLICONS	11
a. <i>Clonage des produits PCR.....</i>	<i>11</i>
b. <i>Transformation des bactéries.....</i>	<i>11</i>
c. <i>PCR sur colonies.....</i>	<i>12</i>
d. <i>Extraction d'ADN plasmidique.....</i>	<i>12</i>
6) SEQUENÇAGE.....	13
7) ANALYSES BIOINFORMATIQUES.....	13
8) QUANTIFICATION DE L'EXPRESSION DES LACCASES	14
a. <i>Extractions d'ARN.....</i>	<i>14</i>
b. <i>Amorces utilisées pour la RT-PCR semi-quantitative et quantitative</i>	<i>14</i>
c. <i>RT-PCR semi-quantitative</i>	<i>15</i>
d. <i>RT-PCR quantitative.....</i>	<i>15</i>
RESULTATS.....	17
1) DES GENES TOUCHES PAR L'INSERTION DE L'ADN-T	17
a. <i>Sites de restriction et amorces spécifiques</i>	<i>17</i>
b. <i>Optimisation des conditions pour la caractérisation des sites d'insertion : utilisation des lignées pBIB HYGRO</i>	<i>20</i>
c. <i>Application de la caractérisation des sites d'insertion sur des lignées pMDC32</i>	<i>28</i>
2) UNE SUR-EXPRESSION DE MIR397A IMPACTE L'EXPRESSION DES LACCASES	31
DISCUSSION	34
BIBLIOGRAPHIE	37
SITOGRAPHIE	39
ANNEXES	40

ABRÉVIATIONS

ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
ADN-T	ADN de Transfert
ARN	Acide ribonucléique
ARNm	ARN messenger
BET	Bromure d'éthidium
BLAST	Basic Local Alignment Search Tool
CaMV	Cauliflower Mosaic Virus
Ct	Cycle threshold (Cycle seuil)
DEPC	Diethyl Pyrocarbonate
dNTP	Désoxyribonucléotides triphosphates
ddNTP	Didésoxyribonucléotides triphosphates
EDTA	Ethylenediaminetetraacetic Acid (acide éthylène diamine tétraacétique)
miARN	Micro ARN
NTP	Nucléotides triphosphates
PCR	Polymerase Chain Reaction (Réaction de Polymérisation en Chaîne)
qPCR	PCR quantitative
rcf	Relative centrifugal force (force centrifuge relative)
rpm	Rotation par minute
RT	Reverse transcription
Taq	Thermus aquaticus
Ti	Tumor inducing (inducteur de tumeur)
Tm	Melting temperature (température de fusion)
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

PRESENTATION SUCCINCTE DE L'ORGANISME D'ACCUEIL

L'INRA, Institut National de Recherche Agronomique, fut créé en 1946 et est à ce jour le 2^{ème} institut de recherche agronomique en Europe. L'INRA s'occupe de nombreuses missions, autant sur les plans économiques que sociaux et environnementaux. Plus précisément, les unités de recherches d'Orléans faisant partie du centre de recherche Val de Loire se préoccupent de domaines de recherche concernant la sélection d'arbres forestiers, la gestion des ressources génétiques de ces espèces, la biologie des insectes forestiers ravageurs, ainsi que la maîtrise des érosions et des pollutions. Mon unité d'accueil, l'AGPF (Amélioration Génétique et Physiologie Forestières) (*voir ANNEXE I*), s'intéresse à la génétique et la physiologie du bois, dans le but de réaliser des sélections génétiques et d'améliorer la production de biomasse.

INTRODUCTION

1) Introduction générale

Actuellement, un des sujets de préoccupation concerne l'énergie : son besoin croissant a conduit au développement de biocarburants de seconde génération. Du bioéthanol peut être obtenu en faisant subir des traitements enzymatiques à des sucres fermentescibles. Cependant, cette conversion en biocarburants est défavorisée par la présence des lignines, un des composants du bois. C'est ce qui a conduit le projet PLANT-KBBE 2010 « TreeForJoules » à s'intéresser à l'amélioration des propriétés du bois. Ce projet utilise deux espèces d'arbre, l'eucalyptus et le peuplier, qui présentent de bonne aptitude à la production de biomasse. Le peuplier, en plus d'avoir une croissance rapide et d'être une espèce à forte production de biomasse, est le premier arbre à avoir été séquencé, rejoignant ainsi les génomes de plantes déjà séquencés, tels que *Arabidopsis thaliana* et le riz (Tuskan *et al*, 2006 ; The Arabidopsis Genome Initiative, 2000 ; Goff *et al*, 2002). Grâce aux différents outils disponibles, les gènes contrôlant les propriétés du bois peuvent être identifiés et étudiés. Le but est d'identifier des gènes importants pour le contrôle de la quantité et/ou de la qualité des lignines, afin d'utiliser ces derniers pour développer des stratégies de sélection assistées par des marqueurs moléculaires ; ceci est un des objectifs du projet PLANT-KBBE 2010 « TreeForJoules ». La transgène est utilisée ici comme outil d'analyse pour étudier la fonction *in planta* de gènes candidats sélectionnés. Dans le cadre de mon stage, j'ai développé une méthode permettant de caractériser les sites d'insertion des transgènes. Cette analyse nous permet de déterminer

l'emplacement de l'insertion de l'ADN transféré dans le génome du peuplier, et donc de sélectionner les lignées transgéniques pour lesquelles nous n'avons pas d'insertion dans un gène, ce qui pourrait induire des effets de mutation insertionnelle.

J'ai également analysé des peupliers transgéniques sur-exprimant un gène codant pour un micro ARN : miR397a. Les cibles prédites de ce micro ARN étant des gènes impliqués dans la déposition de lignines, l'objectif de cette approche est de déterminer si des modulations d'expression de miR397a impactent l'expression des gènes de laccases, et par conséquent la déposition des lignines.

2) Etude bibliographique

Cette partie bibliographique est constituée de deux parties : la première présente les méthodes employées pour caractériser des sites d'insertion de transgène, et la seconde aborde les laccases, leur implication dans la lignification, et le rôle possible de miR397a dans la régulation transcriptionnelle de ces gènes.

a. Les sites d'insertion et leur emplacement

Agrobacterium tumefaciens est une bactérie phytopathogène du sol qui a la capacité d'infecter naturellement un grand nombre de dicotylédones, c'est pourquoi elle est souvent utilisée comme vecteur de transformation génétique chez les plantes. L'infection s'effectue grâce à une partie du plasmide Ti que l'on nomme l'ADN-T (pour ADN de transfert, *voir Figure 1* ; Riva *et al*, 1998). L'ADN-T comprend deux bordures : les bordures gauche et droite. Ce sont ces bordures qui vont être ciblées par des gènes de virulence et permettre l'insertion de l'ADN-T dans le génome de la plante (Gelvin *et al*, 2003). Un gène étranger peut donc être introduit dans ce plasmide entre les deux bordures de l'ADN-T, par exemple avec la technique de clonage Gateway®, afin de produire des cellules végétales génétiquement modifiées (exemple chez *Arabidopsis* : Curtis *et al*, 2002). La position de ces transgènes dans le génome d'un organisme est nommée site d'insertion. La caractérisation des sites d'insertion est importante afin de savoir si l'ADN-T s'est placé dans un gène ou non, son emplacement pourrait ainsi expliquer une croissance particulière des plants, mettant en avant une corrélation entre gène affecté et phénotype : on parle de mutation insertionnelle. Pour cela, il faut déterminer combien il y a de sites d'insertion, où il(s) a (ont) eu lieu et si il(s)-ci s'est (se sont) effectué(es) dans un gène. Pour l'identification de ces sites d'insertion, la stratégie consiste à cloner et séquencer l'ADN génomique flanquant les bordures de l'ADN-T. Plusieurs méthodes ont été développées, notamment la TAIL-PCR (Liu et Whittier, 1995), ou

encore la PCR walking (Devic *et al*, 1997). La première méthode, ne nécessitant pas de préparation particulière de l'ADN au préalable, se base sur trois amplifications. Elle requiert l'utilisation de trois amorces emboîtées (ou « nested ») spécifiques d'un côté, et d'une petite amorce dégénérée de l'autre (Liu et Whittier, 1995) ; elle fut adaptée pour identifier les séquences flanquant les sites d'insertion de l'ADN-T d'*A. tumefaciens* (Liu *et al*, 1995). La technique dite de « PCR walking », elle, se fait à partir d'ADN génomique digéré par des enzymes de restriction ayant des coupures franches. Des adaptateurs sont ensuite ligués sur les fragments d'ADN : trois réactions PCR sont alors mises en place avec des amorces nested, spécifiques de l'ADN-T et de l'adaptateur (Devic *et al*, 1997). La méthode de « PCR walking » a été adaptée pour être utilisée à haut débit, notamment pour la caractérisation des banques de mutants d'insertion d'*A thaliana*. (Balzergue *et al*, 2001). Par exemple, une optimisation a été faite sur les étapes de digestion et ligation des adaptateurs, qui ont été réunies en une seule étape.

Pendant ces recherches des sites d'insertion de l'ADN-T, certaines particularités ont pu être notées, que ce soit chez *Arabidopsis thaliana* ou chez d'autres plantes : des délétions d'ADN génomique lors de cette insertion, que ce soit au niveau de la bordure gauche ou droite de l'ADN-T, ont effectivement été mises en évidence (Kumar *et al*, 2002), mais il a aussi été noté des insertions complexes (Ortega *et al*, 2002).

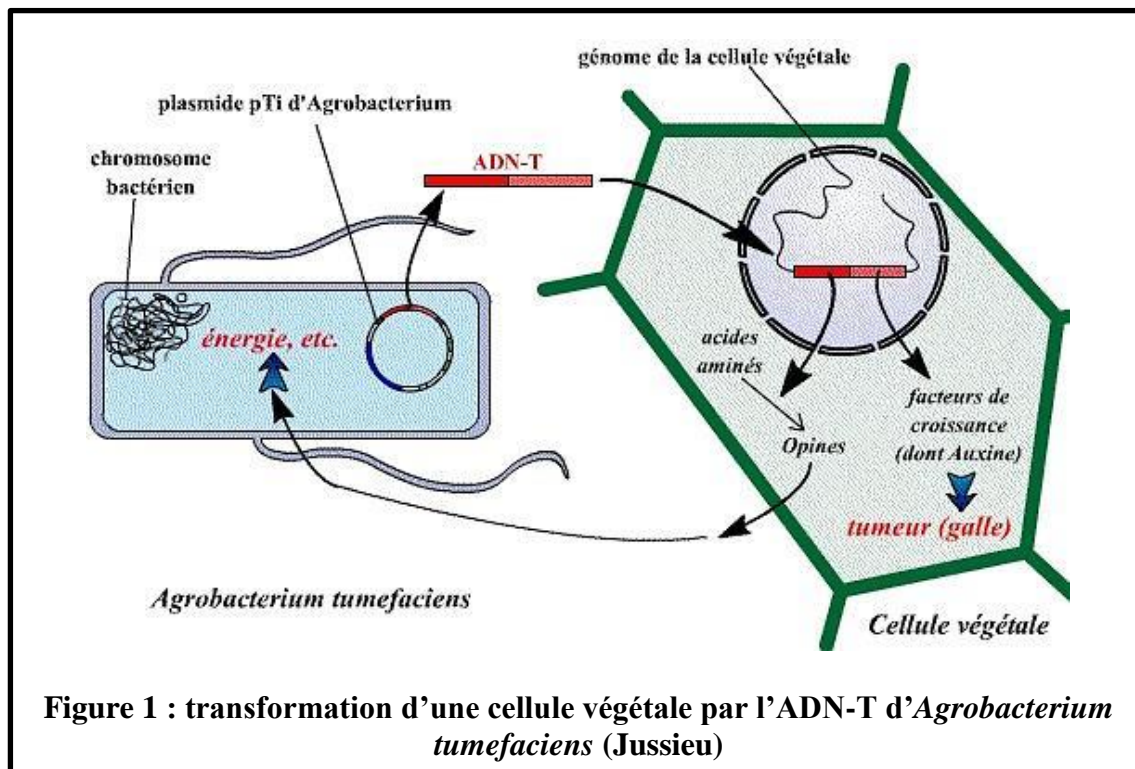


Figure 1 : transformation d'une cellule végétale par l'ADN-T d'*Agrobacterium tumefaciens* (Jussieu)

b. Micro-ARN, laccases et lignification

Le bois, qui constitue la majeure partie de la biomasse végétale, est un tissu ayant une structure complexe résultant d'une succession de plusieurs couches cellulaires comprenant les rayons, les vaisseaux et les fibres (Leplé *et al*, 2004 ; Déjardin *et al*, 2010 ; Plomion *et al*, 2001). Le bois est constitué principalement de cellules mortes (fibres et vaisseaux) possédant une paroi secondaire épaisse. Les parois secondaires se composent principalement de cellulose (40-80 %), d'hémicellulose (10-40 %) et de lignine (5-25 %) (Bidlack *et al*, 1992).

La lignine est essentielle à l'imperméabilisation des vaisseaux permettant la conduction de la sève et, associée à la cellulose, elle confère aux cellules de fibres les propriétés mécaniques nécessaires au soutien mécanique de l'arbre. C'est un polymère amorphe et hydrophobe provenant de la polymérisation de monolignols : alcool *p*-coumarylique, alcool coniférylique et alcool sinapylique (Plomion *et al*, 2001 ; Simmons *et al*, 2010). Ces monolignols ont des proportions différentes en fonction des organismes et de nombreuses enzymes sont impliquées dans leur biosynthèse (Boerjan *et al*, 2003 ; Simmons *et al*, 2010 ; Vanholme *et al*, 2010).

Des études sur la fonction *in planta* des gènes de la lignification ont été menées sur différentes espèces et notamment sur le peuplier (Boerjan *et al*, 2003 ; Pilate *et al*, 2012). Des essais aux champs d'arbres transgéniques modifiés pour l'expression de ces gènes ont été réalisés depuis plus de 20 ans. Ces travaux ont permis de mieux comprendre la fonction de chaque gène dans la voie de biosynthèse, mais ils ont également permis de tester des applications possibles, visant à réduire ou modifier le contenu en lignines (Pilate *et al*, 2012). Si une réduction de la lignine a pu dans certains cas être observée, elle était cependant accompagnée d'effets négatifs sur la croissance.

Pour tenter de diminuer ces effets néfastes, d'autres gènes impliqués dans la voie de biosynthèse des lignines pourraient être ciblés. Après avoir été biosynthétisés, les monolignols sont pris en charge par des peroxydases et des laccases qui assurent la polymérisation oxydative des monolignols. Christensen *et al*. (2001) ont montré que certaines peroxydases étaient impliquées dans la lignification mais qu'elles n'agiraient que tardivement au cours de la différenciation cellulaire (Christensen *et al*, 2001).

Les laccases sont des protéines à cuivre appartenant à une famille multigénique : elles fonctionneraient durant les premières étapes de lignification en polymérisant les monolignols (Berthet *et al*, 2012). Des différences phénotypiques ont pu être observées chez des mutants *Arabidopsis*, sous-exprimant les laccases. Effectivement, des mutants d'insertion sur les gènes *AtLAC4* et *AtLAC17* montrent des modifications au niveau de la composition de la lignine,

ainsi que des altérations du phénotype des tiges (Berthet *et al*, 2011). Des peupliers, modifiés pour l'expression de trois laccases de peuplier (*LAC3*, *LAC90*, et *LAC110*), n'ont cependant pas permis d'observer les mêmes résultats que pour *Arabidopsis*. Les peupliers transformés, et présentant une importante diminution de l'expression d'une laccase (*LAC3*), ne révèlent aucun changement au niveau de la composition de la lignine (Ranocha *et al*, 2002).

Les laccases étant une famille multigénique, il se peut qu'il existe une certaine redondance fonctionnelle et que, par conséquent, le fait d'éteindre l'expression d'une laccase ne suffise pas pour observer une diminution de la lignification (Ranocha *et al*, 2002). Une alternative consiste à éteindre l'expression de plusieurs gènes de laccases en même temps, et si possible les gènes exprimés lors de la lignification. Il peut par exemple être envisagé de modifier l'expression de certains micro-ARN, qui pourraient être des régulateurs transcriptionnels de ces laccases. Les miARN sont capables de reconnaître une séquence complémentaire au niveau de l'ARNm, puis de cliver les ARNm ou d'en réduire la traduction. Il serait donc possible de bloquer l'expression des laccases grâce aux micro-ARN. Notamment, miR397a serait capable de cibler et diminuer l'expression de laccases, telles que *AtLAC4* et *AtLAC17* chez *Arabidopsis* (miRBase ; Abdel-Ghany et Pilon, 2008). Les homologues de ces laccases *AtLAC4* et *AtLAC17* observées chez *Arabidopsis* ont été trouvées chez le peuplier (Berthet *et al*, 2012). Il est à noter que toutes les laccases de peuplier se groupant dans les mêmes clades que les laccases d'*Arabidopsis* impliquées dans la lignification, sont prédites pour être des cibles de miR397a. Ainsi, la sur-expression de miR397a pourrait permettre de réduire l'expression de presque toutes les laccases de peuplier potentiellement impliquées dans la lignification du bois.

3) Présentation du thème du stage

Mon travail de stage s'est déroulé en deux phases indépendantes :

a. La mise au point d'une méthode permettant de déterminer les sites d'insertion de transgènes

La mise au point de la technique pour caractériser les sites d'insertion s'est effectuée en utilisant des lignées transgéniques connues du laboratoire. Les fragments d'intérêt, situés au niveau de chaque extrémité de l'ADN-T inséré dans le génome, sont récupérés grâce à des PCR nichées (ou nested) : la technique est dérivée des travaux de Balzergue *et al*, 2001. Le principe consiste à trouver deux couples d'amorces spécifiques, qui vont permettre

d'amplifier des fragments (situés de chaque côté des bordures gauche et droite de l'ADN-T) préalablement isolés grâce à des enzymes de restriction. Ces fragments sont ensuite séquencés et analysés par bioinformatique : la position de la séquence sur le génome de *Populus trichocarpa* peut alors être déterminée.

b. L'étude d'expression de peupliers transgéniques sur-exprimant le micro ARN miR397a, et l'impact sur l'expression de gènes de laccases.

Les travaux d'analyse fonctionnelle menés sur les laccases d'Arabidopsis et de peuplier, ainsi que les prédictions de régulation de ces laccases par miR397a, indiquent que la modulation d'expression de miR397a pourrait impacter l'expression des laccases impliquées dans la lignification. Afin de vérifier cette hypothèse, des peupliers transgéniques sur-exprimant miR397a sont utilisés pour analyser l'expression des gènes codant les laccases ciblées par ce micro-ARN. Pour cette étude, certaines laccases sont sélectionnées d'après les travaux de Berthet *et al*, 2012. Des mutants Arabidopsis pour *AtLAC4* et *AtLAC17* ont révélés des altérations au niveau de la lignification (Berthet *et al*, 2012). Plusieurs laccases de peuplier, homologues à *AtLAC4* et *AtLAC17* d'Arabidopsis, ont donc été retenues afin d'être étudiées : ce sont les laccases *PtrLAC1*, *PtrLAC15*, *PtrLAC16* et *PtrLAC42*, toutes ciblées par miR397a. Cette analyse est réalisée dans un premier temps avec la méthode de RT-PCR semi-quantitative, permettant d'observer visuellement s'il y a des différences d'intensité dans l'expression de ces gènes de laccases. Ensuite, certaines lignées sont sélectionnées d'après ces observations afin d'effectuer une quantification par RT-PCR quantitative.

MATERIELS ET METHODES

1) Constructions utilisées

Deux types de vecteurs binaires ont été utilisés pour la transformation des peupliers (*Populus tremula* x *Populus alba*).

Le premier vecteur binaire utilisé est le vecteur pBIB-HYGRO (*voir ANNEXE II*), contenant le gène de l'hygromycine phosphotransférase. Ce dernier remplace le gène de la néomycine phosphotransférase trouvé à l'origine dans le vecteur pBIN19 (Becker, 1990).

Le second vecteur binaire utilisé est pMDC32 (*voir ANNEXE II*) et contient quant à lui deux promoteurs CaMV 35S, en plus de gènes de résistance à la kanamycine et à l'hygromycine.

2) Matériel végétal

La lignée non transformée correspond au clone INRA N°717-1B4 (*Populus tremula* x *Populus alba*) utilisé pour la transformation génétique. Les peupliers transgéniques utilisés dérivent des deux types de vecteurs binaires pBIB HYGRO et pMDC32.

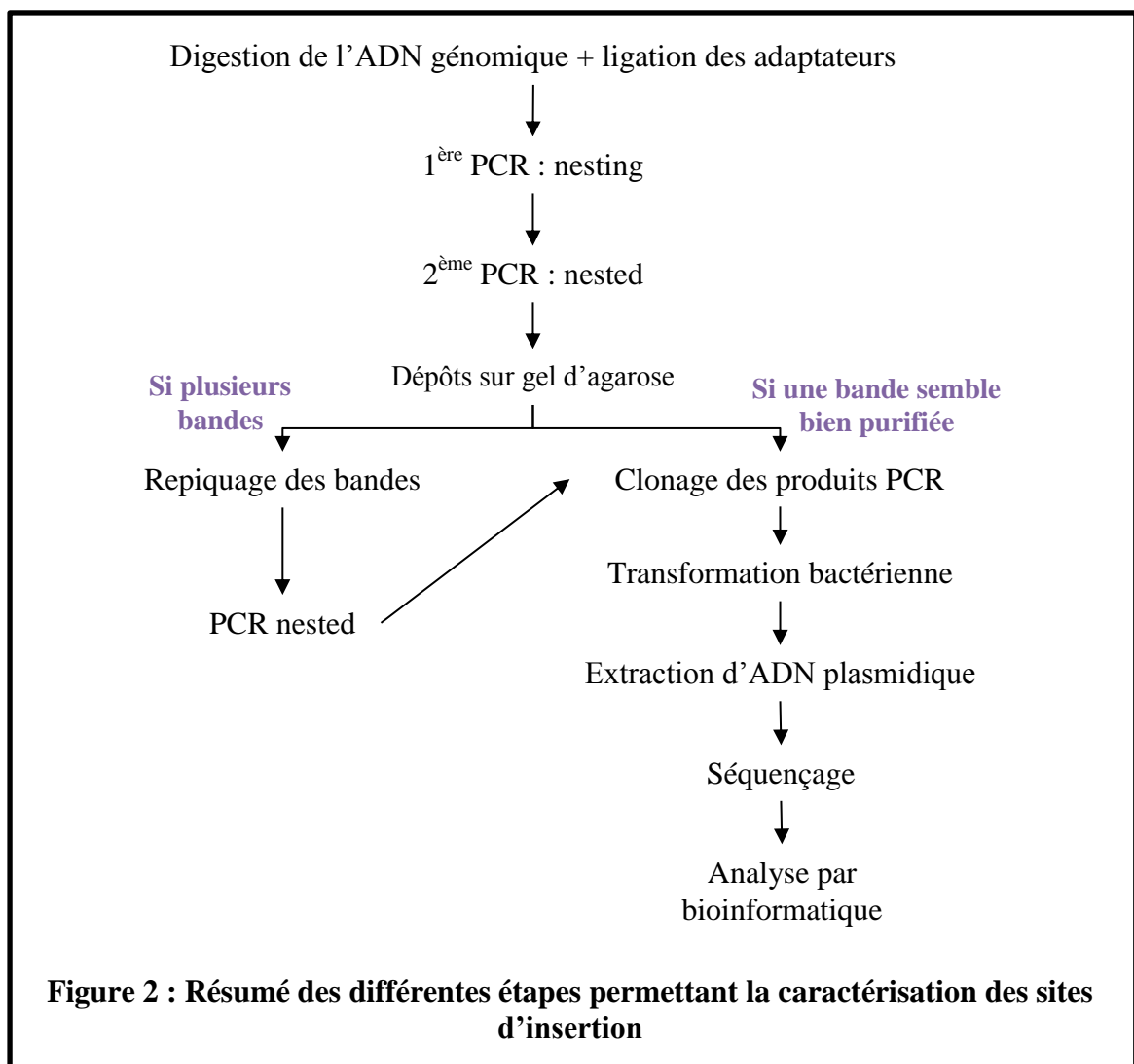
- Identifiants des lignées transformées avec pBIB-HYGRO comme vecteur binaire : FS40, FS3 et FAS13
- Identifiants des lignées transformées avec pMDC32 comme vecteur binaire : Myb156-10d, Myb156-15d, Myb90-19d, Myb90-23d et 556-10
- Lignées transgéniques sur-exprimant miR397a, transformées avec pMDC32 : 35SmIR397a A1 à A35. Pour chaque lignée transformée, trois plants ont été échantillonnés et constituent des répétitions biologiques.

3) Généralités sur la caractérisation des sites d'insertion

La caractérisation de sites d'insertion nécessite plusieurs étapes, représentées schématiquement ci-dessous (*voir Figure 2*). Chacune de ces étapes est détaillée par la suite, les parties correspondantes sont indiquées entre parenthèses.

Des amorces spécifiques du vecteur binaire utilisé pour la transformation transgénique (pBIB HYGRO ou pMDC32), ainsi que les enzymes de restriction à utiliser, doivent dans un premier temps être déterminées. Une fois cette étape accomplie, l'ADN génomique des différents plants transgéniques peut être extrait (§ 4.a). Cet ADN est ensuite digéré par les enzymes de restriction précédemment choisies, puis les adaptateurs (Balzergue *et al*, 2001) sont ligués (§ 4.b). Grâce à cette dernière étape, des PCR peuvent alors être réalisées avec les amorces

spécifiques des adaptateurs et celles spécifiques du vecteur binaire. Ces produits PCR sont ensuite migrés sur gel (§ 4.c) pour vérifier la présence d'amplicons et déterminer leur(s) taille(s). Ces derniers sont alors clonés par T/A cloning en utilisant le plasmide pCR™2.1-TOPO (§ 5.a). Les plasmides sont ensuite transfectés dans des bactéries Top10 (§ 5.b). Après criblage des colonies par PCR, des colonies sont mises en culture, l'ADN plasmidique extrait (§ 5.d), séquencé et analysé (§ 6 et 7).



4) Amplification des sites d'insertion

a. Extractions d'ADN génomique

L'extraction d'ADN génomique est réalisée à partir de feuilles de peuplier provenant de plants cultivés *in vitro*. Les feuilles sont placées dans de l'azote liquide puis broyées grâce à un mortier et un pilon. La poudre obtenue est transférée dans un tube de 2 mL, puis pesée. L'extraction d'ADN génomique est réalisée à partir de 100 mg de poudre en suivant le protocole du kit de Macherey-Nagel, Nucleospin® Plant II. La quantification des acides nucléiques est effectuée grâce à un spectrophotomètre Nanodrop ND1000, avec récupération des données *via* le logiciel ND-1000 V3.1.0.

b. Digestion de l'ADN génomique et ligation des adaptateurs

i. Préparation préalable des adaptateurs :

Le mélange réactionnel est réalisé dans un volume final de 40 μ L. A 38 μ L d'eau sont ajoutés 1 μ L d'ADA1 100 μ M (5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCCGCCGGGG AGGT-3') et 1 μ L d'ADA2 100 μ M (5'-P-ACCTCCCC-N3'). Le mélange est incubé à 94°C pendant 2 min puis ramené de 70°C à 4°C en 1 min 10 s. La solution est ensuite maintenue à 4°C pendant 5 min.

ii. Mix digestion/ligation (selon Balzergue)

Les étapes de digestion et ligation des adaptateurs s'effectuent en même temps (selon Balzergue *et al*, 2001). Celui-ci est composé d'1 μ L d'enzyme(s) de restriction (1U/ μ L), 1 μ L de T4 DNA ligase (1U/ μ L), 4 μ L de tampon ligase 10 X, 4 μ L de PEG 4000, 6 μ L d'adaptateurs préparés au préalable, 10 μ L d'ADN (à environ 50 μ g/L). Le mélange réactionnel est ajusté à 40 μ L avec de l'eau. Ce mélange est incubé la nuit (environ 15h) à température ambiante.

c. PCR et migrations sur gel

Les PCR réalisées sont des PCR nichées, qui se pratiquent avec 2 couples d'amorces. Ainsi, deux réactions vont se suivre : une première PCR (ou PCR nesting) puis une seconde PCR (ou PCR nested). Les amplicons de la PCR nesting dilués 200 fois, 2000 fois ou 20000 fois servent ensuite de matrice à la PCR nested.

Dans le cas de la première PCR, 1 μ L de produit de digestion est utilisé en tant que matrice. Pour la deuxième PCR, 1 μ L des produits dilués de la première PCR est utilisé. Ensuite sont ajoutés 2,5 μ L de tampon PCR 10X, 0,75 μ L de MgCl₂ à 50mM, 0,5 μ L de dNTP à 10 mM, 0,5 μ L de chacune des amorces à 10 μ M et 0,2 μ L de Taq Polymérase (5U/ μ L ; Invitrogen). Le

volume est ajusté à 25 µL avec de l'eau. Les cycles PCR utilisés lors de la caractérisation des sites d'insertion se déroulent comme suit. Une étape préalable consiste à préchauffer à 80°C et mettre les mix PCR à ce moment afin d'éviter qu'il y ait une activité de la polymérase lors de l'augmentation progressive de la température. Puis, suite à une étape de 2 min à 94°C, l'amplification poursuit sur 40 cycles avec pour conditions 94°C pendant 30 s, 67°C pendant 45 s et 72°C pendant 2,5 min. L'étape finale d'élongation s'effectue pendant 5 min à 72°C. L'amplification est réalisée avec un thermocycleur GeneAmp® PCR System 9700 (Applied Biosystems®). 10 µL de chaque produit PCR sont déposés sur gel d'agarose 1,5 % (p : v) (*voir ANNEXE III*) puis révélés sous rayons UV, après avoir été exposés au BET.

Dans le cas où des bandes sont visibles, cela signifie que nos fragments d'intérêt ont bien été amplifiés, et que la caractérisation des sites d'insertion peut continuer avec ces mêmes produits PCR.

5) Clonage et séquençage des amplicons

a. Clonage des produits PCR

Les fragments d'intérêts sont clonés par TOPO cloning reaction, ou TA cloning grâce à un vecteur de clonage nommé pCRTM2.1-TOPO (*voir ANNEXE IV*). Ce vecteur, comprend une désoxythymidine « débordante » en 3', permettant la ligation des produits PCR de manière plus efficace. Le mélange réactionnel utilisé pour ce clonage se compose de 3 µL de produits PCR, 1 µL de solution saline (NaCl 1,2 M ; MgCl₂ 0,06 M), 1 µL d'eau et 1 µL du vecteur pCRTM2.1 TOPO (produits provenant du kit TOPO® TA Cloning® Kit with pCR®2.1-TOPO® Vector). La réaction est incubée 5 min à température ambiante puis placée dans la glace.

b. Transformation des bactéries

Les bactéries utilisées pour les transformations sont des *Escherichia coli* chimiquement compétentes, les Top10 « One Shot® TOP10 Chemically Competent *E. coli* » (InvitrogenTM). Elles permettent d'identifier les bactéries transformées grâce à leur coloration. En effet, le vecteur de clonage pCRTM2.1 TOPO contient le gène lac Z qui code la β-Galactosidase (*voir ANNEXE IV*). Une coloration bleue est induite lorsque la β-Galactosidase hydrolyse le X-Gal. Dans notre cas, ce sont les colonies blanches qui nous intéressent étant donné qu'elles indiquent qu'il y a incorporation d'ADN au niveau du gène lacZ. Cette incorporation inactive ce gène, et rend impossible l'hydrolyse du X-Gal.

Des tubes de bactéries Top10 chimiquement compétentes sont disposés dans de la glace. 2 μL du mélange de clonage obtenu précédemment sont ajoutés aux bactéries. Le tout est ensuite laissé 30 min dans la glace, incubé 30 s à 42°C afin de faire subir un choc thermique, et remis 2 min dans la glace. 250 μL de milieu SOC (tryptone 2%, extrait de levure 0,5%, NaCl 10 mM, KCl 2,5 mM, MgCl_2 10 mM, MgSO_4 10 mM, glucose 20mM ; Invitrogen) sont ajoutés et les tubes sont alors placés sous agitation à 400 rpm pendant une heure à 37°C. Des boîtes de Pétri contenant du milieu LB + Kanamycine (50 mg/L) sont préparées, puis recouvertes de 40 μL de X-Gal. À la fin de cette étape, 50 à 100 μL de la préparation sont étalés sur ces boîtes de Pétri. Les boîtes sont incubées à 37°C jusqu'au lendemain. Des colonies blanches et bleues peuvent alors être observées.

c. PCR sur colonies

De 2 à 30 colonies blanches de chacune des constructions sont reprises dans 10 μL d'eau afin d'effectuer une PCR. Celles-ci sont réalisées dans 2,5 μL de tampon 10X, 0,75 μL de MgCl_2 50 mM, 0,5 μL de dNTP 10mM, 0,5 μL de chaque amorce 10 μM , 0,2 μL de Taq polymérase (5 U/ μL ; Invitrogen) et d'eau pour ajuster le volume à 25 μL . Les amorces utilisées pour l'amplification du fragment sont les amorces M13-F (pour forward) et M13-R (pour reverse), situées sur le vecteur pCRTM2.1 TOPO. Suite à une étape de 2 min à 94 °C, l'amplification se poursuit par 40 cycles avec pour conditions 94°C pendant 30 s, 60°C pendant 45 s et 72°C pendant 2,5 min. L'étape finale d'élongation s'effectue pendant 5 min à 72°C. 10 μL des produits PCR sont ensuite analysés sur gel d'agarose 1,5 % (p:v) afin de déterminer quelles colonies, parmi celles sélectionnées, contiennent effectivement un plasmide avec insert.

d. Extraction d'ADN plasmidique

Parmi les colonies déterminées comme étant positives pour la transformation, 2 colonies de chaque condition sont mises en culture dans du milieu LB liquide contenant de la Kanamycine (50 mg/L). Chaque colonie est reprise dans 3 x 1,5 mL de LB + Kanamycine et les cultures sont laissées en agitation pendant la nuit à 400 rpm et à 37°C.

L'extraction d'ADN plasmidique s'effectue selon les instructions fournies par le kit de Macherey-Nagel, Nucleospin® Plasmid. Suite à l'extraction, l'ADN plasmidique est analysé au Nanodrop.

6) Séquençage

La réaction de séquençage s'effectue à partir de 400 ng d'ADN plasmidique, en utilisant le kit BigDye® Terminator v3.1 Cycle Sequencing (Invitrogen). Deux séquençages sont effectués pour un même échantillon : l'un avec l'amorce M13-F, et l'autre avec l'amorce M13-R, ce qui permet d'avoir un séquençage bi-directionnel et donc une séquence avec moins d'erreurs. L'ADN plasmidique est dans un premier temps dénaturé pendant 3 min à 96°C. Le mix PCR est ensuite ajouté : 8 µL d'ADN à 50 µg/µL (environ), 1 µL d'amorce M13 F ou R (3,2 µM), 3 µL de tampon 5X, 2 µL du mix de séquençage (contenant l'enzyme, les dNTPs ainsi que les dNTPs fluorescents) et de l'eau pour ajuster à 20 µL. L'amplification se déroule finalement sur 30 cycles, avec pour conditions 96°C pendant 10 s, 50°C pendant 5 s et 60°C pendant 4 min.

Les produits PCR sont ensuite purifiés sur colonne Sephadex (NucleoSEQ). Les colonnes sont centrifugées rapidement puis 600 µL d'eau ultra pure sont ajoutés. Les colonnes sont laissées au minimum 30 min au repos, dans des tubes de 2 mL fournis. Le bouchon des colonnes est alors enlevé et les colonnes mises à centrifuger 2 min à 750 rcf. Les colonnes sont transférées dans des tubes de 1,5 mL puis les 20 µL de nos produits PCR précédents y sont déposés. Les colonnes sont mises à centrifuger 5 min à 750 rcf et les échantillons sont répartis dans une plaque PCR. Le séquençage est alors effectué par un séquenceur Genetic Analyser 3500, de Applied Biosystems®, qui comprend 8 capillaires. Cet appareil effectue le séquençage selon la méthode de Sanger, grâce aux ddNTP fluorescents.

7) Analyses bioinformatiques

Les séquences sont récupérées grâce au séquenceur Genetic Analyser 3500 (Applied Biosystems®). Ces séquences, ayant été obtenues avec les amorces M13 présentent sur le vecteur pCR2.1TMTOPO, contiennent une partie de celui-ci. Afin d'isoler la séquence d'intérêt, les amorces M13 sont identifiées, et tout ce qui correspond à la partie vecteur est enlevé. Un BLAST est ensuite réalisé sur le génome du peuplier à partir de cette séquence d'intérêt *via* le site Phytozome (www.phytozome.net ; Goodstein *et al.*, 2011).

8) Quantification de l'expression des laccases

a. Extractions d'ARN

L'extraction d'ARN est réalisée à partir de tige de peuplier. Le xylème est broyé dans de l'azote liquide grâce à un mortier et un pilon. La poudre obtenue est transférée dans un tube de 2 mL, puis pesée. Les ARN totaux sont extraits à partir de 100 mg de la poudre obtenue en utilisant le kit RNeasy Mini, de Qiagen. L'ARN extrait est alors dosé par Nanodrop.

b. Amorces utilisées pour la RT-PCR semi-quantitative et quantitative

Les amorces utilisées pour l'étude de l'expression des laccases ont été désignées de façon à pouvoir amplifier les gènes laccases de peuplier *PtrLAC1*, *PtrLAC15*, *PtrLAC42* ou *PtrLAC16* ainsi que leurs paralogues identifiés dans Berthet *et al.* (2012) (*voir Table I*). Ces amorces sont utilisées pour étudier le gène cible par des techniques de RT-PCR semi-quantitative et de qPCR. Concernant les témoins positifs, des amorces ont été utilisées pour amplifier le gène 18S en RT-PCR semi-quantitative ; et pour amplifier le gène de l'ubiquitine (Ubiquitin-conjugating enzyme, *Potri.012G033000*) pour la RT-PCR quantitative (*voir Table II*).

Table I : Amorces utilisées pour l'étude de l'expression des laccases

Nom des amorces	Séquences (5'-3')	Taille des amplicons
PtrLac1F	CTCTGATGCCTACACTATTAATGG	237 pb
PtrLac1R	GTTTGTCCAGGGGCAATGAG	
PtrLac16F	CTCTGATGCTCACACAATTAACGG	237 pb
PtrLac16R	GTCTGGCCTGGGGCAATTAG	
PtrLac15/42F	CACGCACATATTCTCTGGCT	419 pb
PtrLac15/42R	TGGCCTGGGGCAATTAGG	

Table II : Amorces utilisées pour les témoins positifs en RT-PCR

Nom des amorces	Séquences (5'-3')
18S-F	CTTCGGGATCGGAGTAATGA
18S-R	GCGGAGTCCTAGAAGCAACA
UBC146A	CCCGGCTCTAACCATATCCA
UBC146B	GGGTCCAGCTTCTTGCAGTC

c. RT-PCR semi-quantitative

Les ARN totaux extraits subissent une reverse transcription, le mélange se fait dans un volume final de 20 μL . Il doit contenir 2 μg d'ARN purifiés dans 11 μL d'eau, auxquels sont ajoutés 1 μL d'oligo(dT)₁₂₋₁₈ 0,5 $\mu\text{g}/\mu\text{L}$ et 1 μL de dNTPs 10 mM. Ce mélange est incubé pendant 5 min à 65°C, puis refroidi 3 à 5 min sur glace, centrifugé brièvement et incubé 2 min à 42°C. On ajoute ensuite un mix de réaction qui comprend 4 μL de tampon 5X, 2 μL de DTT (Dithiothréitol) 0,1 M et 1 μL de SuperScript II 200 U/ μL . Les échantillons sont ensuite incubés 50 min à 42°C. Pour terminer la réaction, ils sont mis 15 min à 70°C. Pour ces étapes, le thermocycleur PTC-100 Programmable Thermal Controller de MJ research est utilisé.

Les ADNc sont dilués au 1/4, puis 2 μL de ces ADNc dilués sont utilisés pour la réaction de PCR. Ensuite sont ajoutés 0,5 μL de chaque amorce 10 μM , 0,5 μL de dNTPs 10 mM, 0,25 μL de Taq polymerase (5 U/ μL), 0,75 μL de MgCl_2 50 mM et au final 2,5 μL de tampon 10X. Le volume est ajusté à 25 μL avec de l'eau. L'amplification débute par une étape à 95°C pendant 5 min puis poursuit sur 25 cycles avec pour conditions 95°C pendant 30 s, 60°C pendant 45 s et 72°C pendant 45 s. L'étape finale d'élongation s'effectue pendant 10 min à 72°C.

d. RT-PCR quantitative

La réaction de reverse transcription se fait ici dans un tube de 0,5mL. A 1 μg d'ARN purifié sont ajoutés 10 μL de mix RT 2X (oligo(dT)₂₀ 2,5 μM ; hexamères 2,5 ng/ μL ; MgCl_2 10 mM et dNTPs 10 mM) et 2 μL de mix enzyme RT (SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR, Invitrogen). Le volume est finalement ajusté à 20 μL avec de l'eau DEPC. Le mélange réactionnel est laissé 10 min à température ambiante, puis mis 50 min à 42°C dans un thermocycleur MiniCycler™ PTC-150 de MJ Research. Pour terminer la réaction, il est incubé 5 min à 85°C puis mis sur glace. Enfin, 1 μL de RNase H (2 U/ μL) est ajouté, puis le tube est incubé 20 min à 37°C.

La réaction de PCR quantitative se fait dans un volume réactionnel final de 12 μL . A 2 μL d'ADNc dilué ou non sont ajoutés 0,25 μL de chaque amorce 10 μM , 6 μL de Platinum® SYBR® Green qPCR Super-Mix-UDG 2X (Invitrogen) et 3,5 μL d'eau. Un témoin négatif est également réalisé en mettant de l'eau à la place de la matrice. Le cycle PCR programmé se déroule avec l'appareil Realplex² MasterCycler ep gradient S (Eppendorf) comme suit : après une étape de 2 min à 50°C, il y a une première dénaturation pendant 2 min à 95°C ; l'amplification se poursuit ensuite sur 40 cycles avec pour conditions 95°C pendant 15 s et 60°C pendant 30 s. La mesure de la fluorescence est effectuée à chaque fin de cycle et est

proportionnelle à la concentration d'amplicons présents dans l'échantillon. Dans l'objectif de vérifier la spécificité, une courbe de fusion est générée avec une augmentation progressive de la température jusqu'à 95°C pendant 20 min, pour laisser ensuite 15 s à 95°C.

RESULTATS

Dans l'objectif de voir chez les plants transgéniques l'endroit où l'ADN-T a pu s'insérer, la caractérisation des sites d'insertion est essentielle. Grâce à cette procédure, la caractérisation de quelques sites a pu être achevée sur certaines lignées, notamment la lignée FS40. Il a par ailleurs été noté lors de cette étude que certaines insertions se sont produites dans un gène. Pour y arriver, il faut s'intéresser dans un premier temps au vecteur binaire qui a été utilisé pour l'obtention des plants transgéniques. La détermination des amorces spécifiques, qui sont destinées à amplifier le fragment d'intérêt, dépend en effet du type de vecteur binaire utilisé. Ainsi, la caractérisation des sites d'insertion a été mise au point sur deux types de vecteurs binaires : pBIB HYGRO, anciennement utilisé mais pour lequel on connaît à priori le nombre de sites d'insertion, ceci grâce à des Southern effectués auparavant ; et pMDC32, actuellement très utilisé dans l'unité, et qui servira pour la caractérisation ultérieure des lignées pMDC32 sur-exprimant miR397a.

Quant à l'étude de l'expression des laccases, nous avons observé qu'une sur-expression de miR397a conduisait, pour certaines lignées transgéniques, à une réduction d'expression des laccases prédites pour être régulées par miR397a.

1) Des gènes touchés par l'insertion de l'ADN-T

a. Sites de restriction et amorces spécifiques

Plusieurs lignées ont été caractérisées : FS40, FS3 et FAS13 ayant pBIB HYGRO comme vecteur binaire ; et les lignées Myb156-10d, Myb156-15d, Myb90-19d, Myb90-23d et 556-10 avec pMDC32. Elles sont utilisées afin de mettre au point la méthode de caractérisation des sites d'insertion, dans l'objectif de caractériser les lignées pMDC32 sur-exprimant miR397a. Pour chacun de ces vecteurs, des enzymes de restriction et des amorces spécifiques ont été définies.

Les enzymes de restriction doivent répondre à différents critères. Premièrement, elles doivent avoir des coupures franches, ce qui permet d'utiliser un même adaptateur pour plusieurs enzymes. Deuxièmement, elles doivent couper l'ADN génomique en moyenne tous les 500-1000 pb, pour faciliter l'amplification par PCR. Finalement, elles doivent idéalement couper à une distance raisonnable de la bordure droite ou gauche de l'ADN-T (en fonction de la caractérisation qui doit être effectuée) afin de permettre la conception d'amorces spécifiques du vecteur binaire, et également pour des facilités de séquençage.

Les séquences des vecteurs binaires sont connues et utilisées afin de déterminer des enzymes qui correspondraient à ces différents critères. Les enzymes utilisées pour la caractérisation des sites d'insertion des lignées ayant pBIB HYGRO ou pMDC32 en tant que vecteur binaire sont décrites dans la Table III.

L'adaptateur, ainsi que les amorces utilisées pour l'adaptateur, proviennent des travaux de Balzergue *et al*, 2001. L'adaptateur correspond aux deux oligonucléotides ADA1 (5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCGGGGAGGT-3') et ADA2 (5'-ACCTCCCC-N3'). L'oligonucléotide nommé ADA2 possède un groupe amine en 3' afin d'éviter une extension par la polymérase. Les amorces spécifiques de chaque vecteur binaire ont alors été conçues pour que leurs Tm correspondent avec les amorces spécifiques des adaptateurs (*voir Table IV*), et permettre l'amplification des séquences flanquantes inconnues (*voir Figure 3*). Les Tm ont été comparés grâce au calculateur en ligne OligoCalc, de Northwestern University (<http://www.basic.northwestern.edu/biotools/OligoCalc.html>).

Table III : Enzymes de restriction désignées pour la caractérisation des sites d'insertion

Vecteur binaire	Enzymes	Bordure de l'ADN-T à déterminer	Distance entre la coupure et la bordure à déterminer	Moyenne de coupure sur le génome
pBIB	<i>PsiI</i>	Bordure gauche	200 pb	723 pb (en association avec <i>EcoRV</i>) 600 pb
	<i>RsaI</i>	Bordure droite	300 pb	
pMDC32	<i>HincII</i>	Bordure gauche	476 pb	1000 pb (en association avec <i>EcoRV</i>) 260 pb
	<i>HincII</i>	Bordure droite	260 pb	

Table IV : Séquences des amorces

Amorces utilisées pour l'adaptateur (Balzergue <i>et al</i> , 2001)				
Nom de l'oligonucléotide		Séquences (5' - 3')	Longueur (pb)	Tm (°C)
Ap1b		GGATCCTAATACGACTCACTATAGGGC	27	57,98
Ap2b		CTATAGGGCTCGAGCGGC	18	55,64
Amorces pour les PCR spécifiques utilisées pour les vecteurs binaires				
Vecteurs binaires	Nom des amorces	Séquences (5' - 3')	Longueur (pb)	Tm (°C)
pBIB	LB-BIB-1	CAATCAGCCTGTTGCCGTCTCAC	23	59,12
	LB-BIB-2	CACCCAGTACATTA AAAACGTCCG	25	56,65
	RB-BIB-1	GTCAGTTCCAAACGTAAAACGGCTTG	26	58,24
	RB-BIB-2	CATAACGTGACTCCCTTAATTCTCC	25	54,90
pMDC32	LB-MDC32-1	CCTATAGGGTTTCGCTCATGTGTTG	25	57,52
	LB-MDC32-2	CCAGATCCCCCGAATTAATTCG	22	54,26
	RB-MDC32-1	GCGAATGCTAGAGCAGCTTGAG	22	59,99
	RB-MDC32-2	GATTGTCGTTTCCCGCCTTCAG	22	56,72

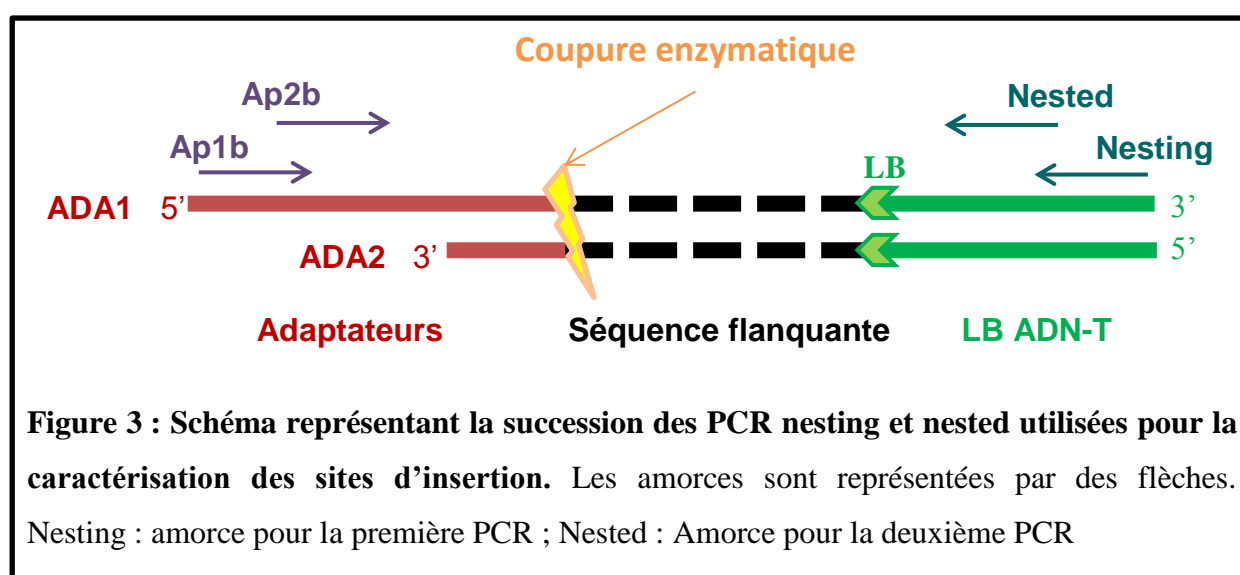


Figure 3 : Schéma représentant la succession des PCR nesting et nested utilisées pour la caractérisation des sites d'insertion. Les amorces sont représentées par des flèches. Nesting : amorce pour la première PCR ; Nested : Amorce pour la deuxième PCR

***b. Optimisation des conditions pour la caractérisation des sites d'insertion :
utilisation des lignées pBIB HYGRO***

Afin de mettre au point les différentes étapes de la méthode, nous avons utilisés 2 lignées transgéniques précédemment caractérisées par la technique de Southern, et donc pour lesquelles nous connaissons le nombre de sites d'insertions : lignées FS40, FS3 et FAS13 avec respectivement 1, 2 et 4 sites d'insertion. Ceci présentait un intérêt pour la mise au point car nous pouvions nous appuyer sur ces données pour voir si la méthode employée était convenable. Nous avons ainsi pu obtenir l'amplification d'un fragment correspondant au nombre de site d'insertion évalué en Southern pour la lignée FS40. Nous avons par ailleurs déterminé que FS40 et FS3 n'étaient en fait qu'une seule et même lignée. Quant à FAS13, un site d'insertion a pu être entièrement caractérisé sur les quatre attendus.

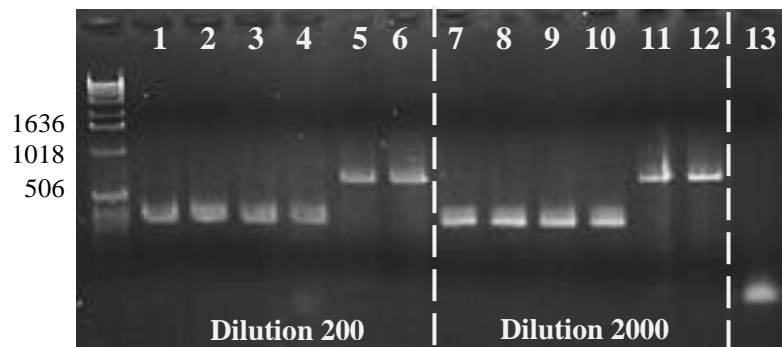
Pour caractériser les sites d'insertion en bordure gauche de pBIB HYGRO, l'ADN génomique de ces lignées est digéré par *Psi I* + *EcoRV* ou *PsiI* seule, tandis que pour la bordure droite *RsaI* est utilisée.

- **Optimisation des conditions pour la caractérisation des sites d'insertion : lignée FS40/FS3**

Une PCR nesting est réalisée sur 2 lots de tubes, avec les amorces Ap1b et LB-BIB-1 ou RB-BIB-1, afin de tester des températures d'appariement différentes. Un premier lot d'échantillons est exposé à une température d'appariement de 67°C et un deuxième lot à une température de 60°C. Ces premières PCR sont alors reprises et diluées 200 fois ou 2000 fois. Pour chaque dilution une deuxième PCR, nommée PCR nested, est réalisée en testant les deux températures d'appariement, en utilisant les amorces Ap2b et LB-BIB-2 ou Ap2b et RB-BIB-2. Cette procédure, utilisant deux couples d'amorces, permet d'amplifier une séquence de manière doublement spécifique, et d'avoir une plus grande sensibilité. Les produits des PCR nested sont déposés sur gel d'agarose à 1,5% (p/v) (*voir Figure 3*). Ces premiers résultats nous montrent des profils semblables pour FS40 et FS3, quelle que soit la dilution de la PCR nesting ou la température d'appariement utilisée. Il faut cependant noter que les bandes sont plus nettes avec une température d'appariement de 67°C et une dilution de la PCR nesting de 2000 fois. Concernant la bordure gauche, une même bande de 350 pb environ est observée, que ce soit une digestion par *PsiI* seule ou *PsiI* + *EcoRV*, et que ce soit la lignée FS40 ou FS3. Pour la bordure droite, une bande de 700 pb environ est visible, pour FS40 et FS3.

(a) 60°C d'appariement

Puits	1	2	3	4	5	6		13
	7	8	9	10	11	12		
Marqueur 1Kb	FS40 <i>PsiI</i>	FS40 <i>PsiI</i> + <i>EcoRV</i>	FS3 <i>PsiI</i>	FS3 <i>PsiI</i> + <i>EcoRV</i>	FS40 <i>RsaI</i>	FS3 <i>RsaI</i>		Témoin négatif (H ₂ O)



(b) 67°C d'appariement

Puits	1		2	3	4	5	6	7
			8	9	10	11	12	13
Marqueur 1Kb	Témoin négatif (H ₂ O)		FS40 <i>PsiI</i>	FS40 <i>PsiI</i> + <i>EcoRV</i>	FS3 <i>PsiI</i>	FS3 <i>PsiI</i> + <i>EcoRV</i>	FS40 <i>RsaI</i>	FS3 <i>RsaI</i>

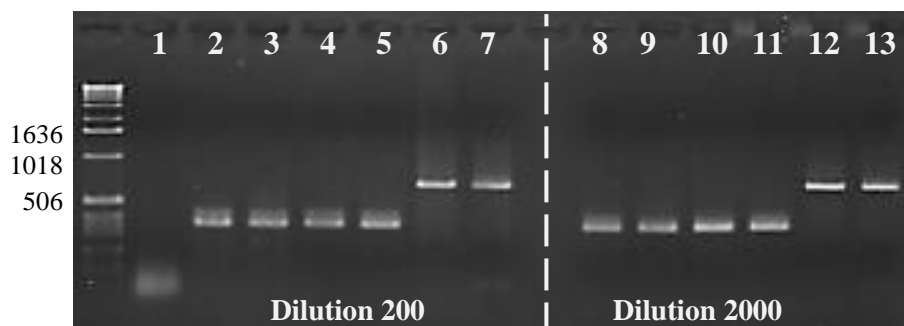


Figure 3 : Séparation sur gels d'agarose 1,5 % des amplicons obtenus par une 2^{ème} PCR « nested » de FS40 et FS3, digérés par *PsiI*, *PsiI* + *EcoRV* ou *RsaI*, avec une température d'appariement de (a) 60°C ou (b) 67°C.

Les bandes étant mieux définies avec une dilution de 2000 de la 1^{ère} PCR, ces produits sont utilisés pour être clonés. Suite au clonage, nous réalisons une PCR sur bactéries afin de vérifier la présence de l'insert. Les amorces M13 ayant été employées, les amplicons ont 200 pb en plus par rapport à la taille du fragment d'intérêt original (*voir ANNEXE IV*), ce qui doit être pris en compte lors de l'analyse des gels. Nous amplifions des bandes de tailles attendues, d'environ 550 pb pour la bordure gauche et 900 pb pour la bordure droite (*voir Figure 4*). Nous avons pu grâce à ces gels déterminer quelles bactéries sélectionnées avaient bien intégrées notre fragment d'intérêt. Parmi ces colonies, seules 2 pour chaque condition de digestion sont récupérées, puis l'ADN plasmidique est extrait.

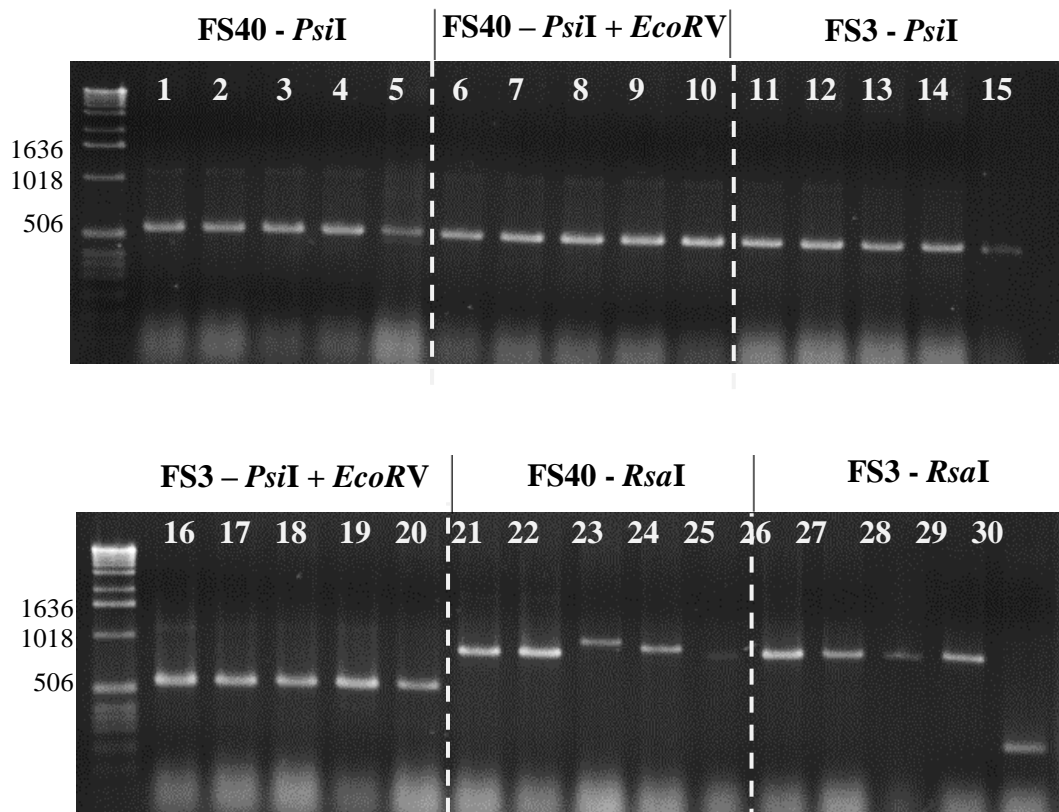


Figure 4 : Migration sur gels d'agarose 1,5% des PCR réalisées sur colonies transformées par des fragments de FS40 ou FS3, qui ont été digérés par *PsiI*, *PsiI* + *EcoRV* ou *RsaI*.

Une fois les séquences récupérées, elles sont nettoyées des parties de vecteur pCR®2.1-TOPO amplifiées par les amorces M13, puis alignées sur le génome du peuplier par BLAST en utilisant les ressources présentes sur le site d'annotation officiel du génome de *Populus trichocarpa* (www.Phytozome.net). Les analyses des séquences de FS40 et FS3 montrent que ces deux lignées ne sont au final qu'une seule et même lignée, étant donné que le site d'insertion retrouvé est le même, et ce que ce soit pour la bordure gauche ou droite. Le site d'insertion du côté de la bordure gauche de l'ADN-T, se trouve dans une zone du génome ne codant pour aucun gène, le chromosome 6, entre les positions 10825685 et 10825438 (voir [Figure 5](#)). La séquence de la bordure droite est aussi positionnée au niveau du chromosome 6, des positions 12066659 à 12066763, et est également dans une zone du génome ne codant pour aucun gène (voir [Figure 6](#)).

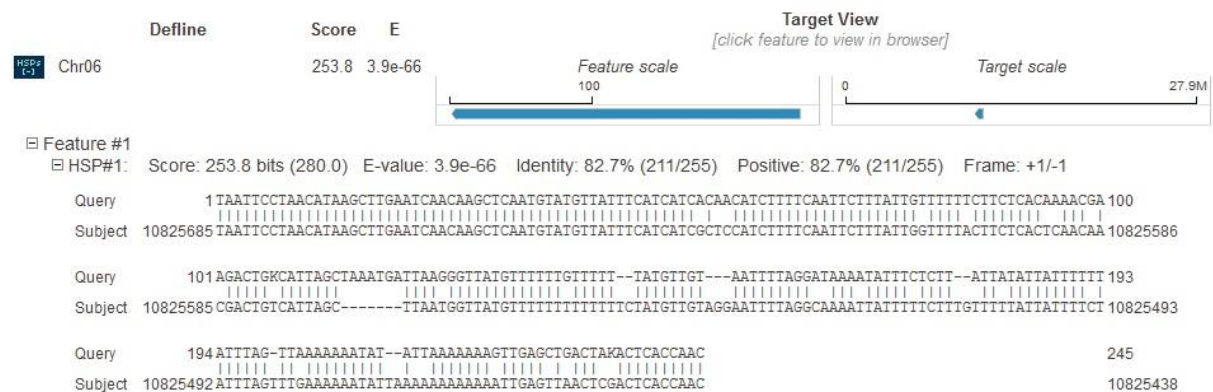


Figure 5 : Détermination d'un site d'insertion au niveau de la bordure gauche d'un ADN-T inséré chez FS40. Copie d'écran après analyse sur Phytozome montrant le résultat du BLAST et sa position sur le chromosome 6 de *Populus trichocarpa*.

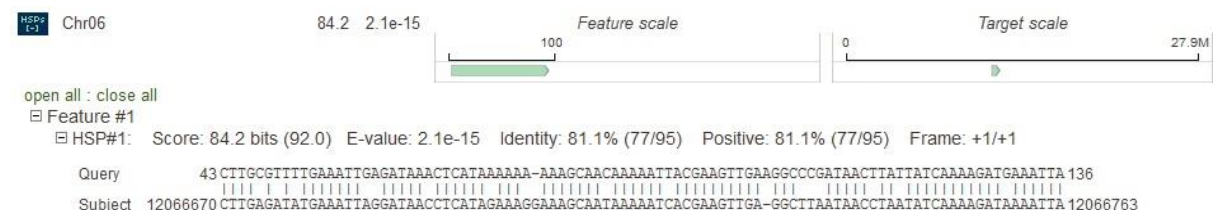


Figure 6 : Détermination d'un site d'insertion au niveau de la bordure droite d'un ADN-T inséré chez FS40. Copie d'écran après analyse sur Phytozome montrant le résultat du BLAST et sa position sur le chromosome 6 de *Populus trichocarpa*.

- **Optimisation des conditions pour la caractérisation des sites d'insertion : lignée FAS13**

La lignée FAS13 doit comporter 4 sites d'insertions d'après les analyses de Southern réalisées auparavant. En utilisant les mêmes conditions optimisées pour FS40/FS3 (température d'appariement de 67°C et première PCR diluée 2000 fois), nous obtenons en deuxième PCR l'amplification de plusieurs fragments. Pour FAS13 LB des bandes à environ 750 pb et 630 pb sont visibles, alors que pour FAS13 RB, on observe des amplicons de 730 pb, 510 pb, 290 pb et 250 pb environ (résultats non montrés).

De même que pour FS40/FS3, nous avons tenté de cloner les différents amplicons visibles. Un nombre important de colonies a donc été criblé et séquencé (*Figures 7 et 8*). Une PCR sur une trentaine de colonies est réalisée pour FAS13 LB et FAS13 RB avec les amorces M13-F et M13-R, afin de cribler les bactéries ayant intégré un fragment d'intérêt. Sur gel, des bandes de tailles différentes sont visibles : nous avons des bandes de 450 et 850 pb pour FAS13 LB ; et de 450, 500, 750 et 900 pb environ pour FAS13 RB (*voir Figures 5 et 6*). Du fait de la position des amorces M13 sur le vecteur pCR2.1 TOPO (*voir ANNEXE IV*) 200 pb sont ajoutées au fragment d'intérêt cloné. Nous analysons (si possible) 2 colonies pour chaque taille différente, si c'est possible (*voir Figures 5 et 6*).

Il est à noter que l'emplacement d'un seul un site d'insertion a pu être caractérisé lors de cette étude. Celui-ci se trouve, comme pour le site caractérisé pour la lignée FS40, sur le chromosome 6 (*voir Figure 9* pour la bordure gauche, et *Figure 10* pour la bordure droite), la bordure gauche entre les positions 10594261 et 10594403 et la bordure droite entre 10593813 et 10594243. Contrairement à la lignée FS40, le site d'insertion se trouve au niveau d'un gène, *Potri.006G129600*, annoté sur Phytozome comme étant un gène d'ubiquitine.

Les autres séquences ayant été analysées n'ont pu aboutir à l'identification de l'emplacement d'autres sites d'insertion. Pour FAS13 LB une seule séquence a finalement été obtenue : les bandes de tailles différentes sur gel étaient dues à la présence d'une partie du vecteur pBIB HYGRO. Quant à FAS13 RB, les autres séquences s'alignaient à de multiples endroits et ne pouvaient donc pas être identifiées (résultats non montrés).

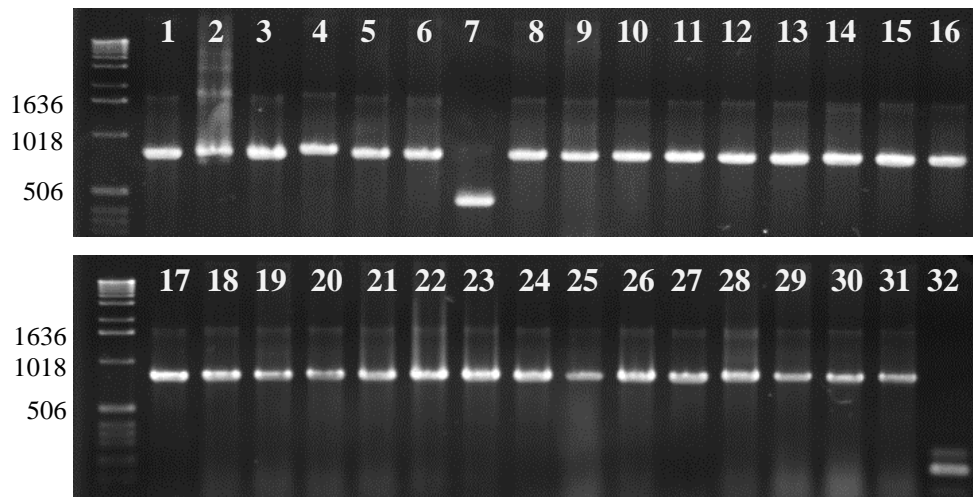


Figure 7 : Migration sur gels d'agarose 1,5% des PCR réalisées sur 31 colonies blanches de FAS13 LB, ainsi qu'un témoin négatif sur colonie bleue en dernière position

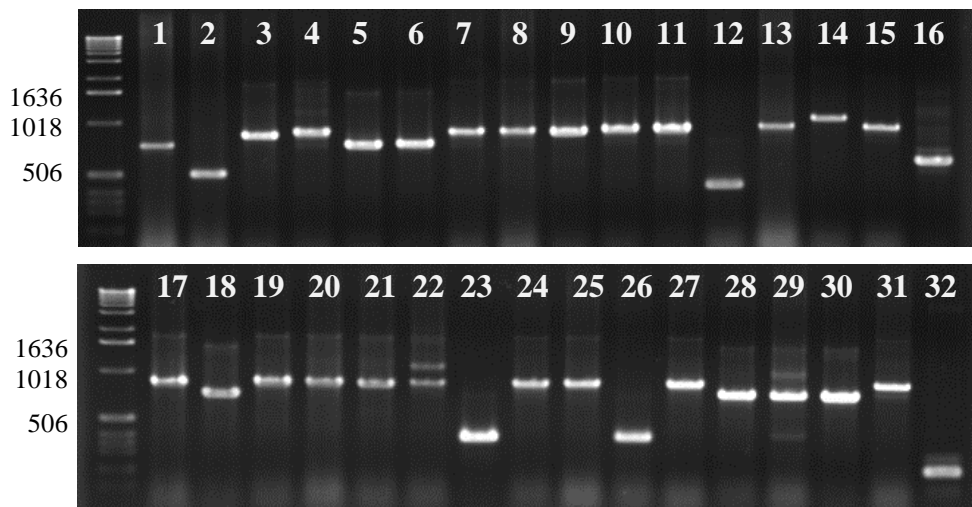
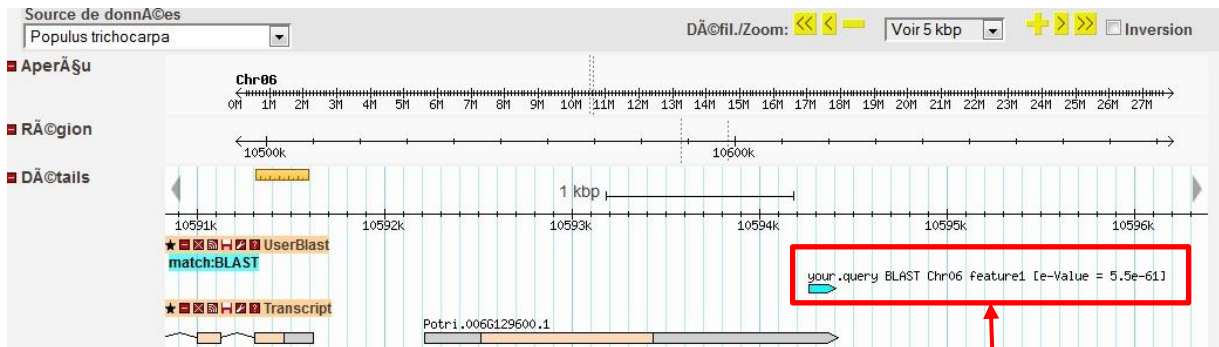
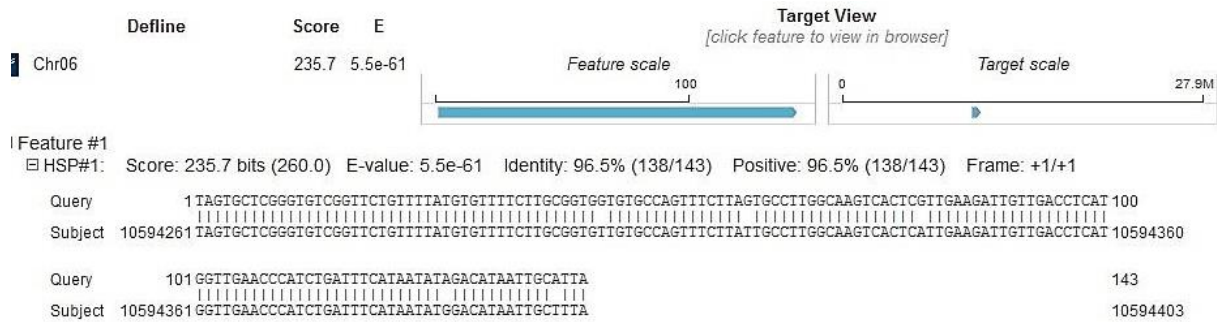


Figure 8 : Migration sur gels d'agarose 1,5% des PCR réalisées sur 31 colonies blanches de FAS13 RB, ainsi qu'un témoin négatif sur colonie bleue en dernière position



Position de
notre sÃ©quence

Figure 9 : DÃ©termination d'un site d'insertion au niveau de la bordure gauche d'un ADN-T insÃ©rÃ© chez FAS13. Copies d'Ã©cran aprÃ©s analyse sur Phytozome montrant le rÃ©sultat du BLAST et sa position sur le chromosome 6 de *Populus trichocarpa* au niveau du gÃ©ne *Potri.006G129600.1* (Phytozome)

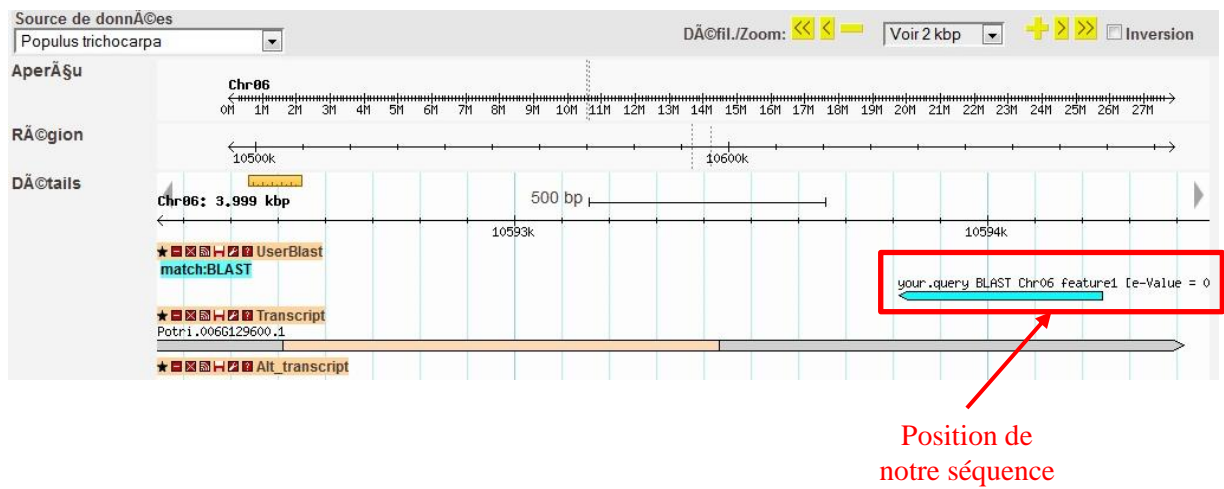
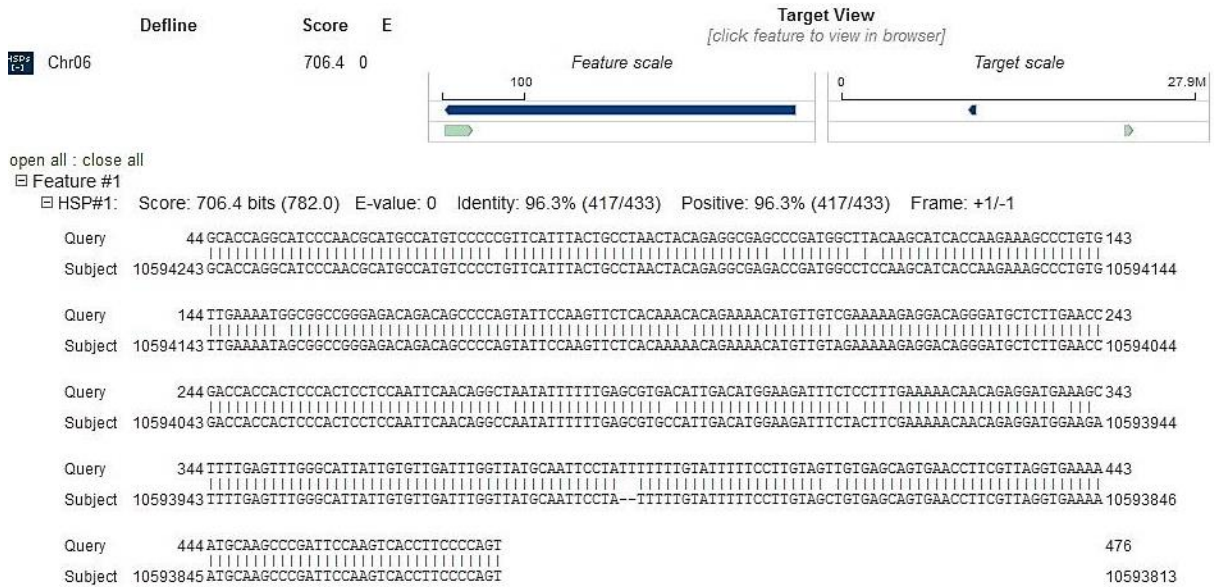


Figure 10 : DÃ©termination d'un site d'insertion au niveau de la bordure droite d'un ADN-T insÃ©rÃ© chez FAS13. Copies d'Ã©cran aprÃ©s analyse sur Phytozome montrant le rÃ©sultat du BLAST et sa position sur le chromosome 6 de *Populus trichocarpa* au niveau du gÃ©ne Potri.006G129600 (Phytozome)

c. Application de la caractérisation des sites d'insertion sur des lignées pMDC32

La méthode de détermination des sites d'insertions, mise en place sur des lignées transformées avec le vecteur pBIB HYGRO, a également été utilisée sur des lignées transformées avec le vecteur pMDC32 : Myb156-10d, Myb156-15d, Myb90-19d, Myb90-23d et 556-10 renommées ici pour plus de simplicité respectivement 10d, 15d, 19d, 23d et 556-10. Le nombre d'insertion est compris entre 1 et 3 en fonction des lignées et avait été déterminé auparavant par PCR quantitative, méthode moins précise que le Southern. Pour la lignée 19d, nous avons réussi à identifier deux sites d'insertion mais au niveau d'une seule bordure pour chaque site.

Pour caractériser les bordures gauche et droite de pMDC32, l'ADN génomique de ces lignées est digéré par *HincII*. Des fragments sont amplifiés pour 19d LB, 556-10 LB et 19d RB seulement : deux bandes aux alentours de 450 pb et 150 pb pour 19d LB ; une bande principale de 600 pb pour 556-10 LB ; et 3 bandes de 750, 350 et 250 pb pour 19d RB (*voir Figure 11*). La PCR nesting est réalisée avec les deux températures d'appariement employées précédemment (60°C et 67°C), puis les produits de PCR nesting sont dilués 2000 fois. Deux lots de tubes sont préparés pour la PCR nested, destinés à une température d'appariement de 60°C ou 67°C. Les gels ne montrent des amplifications que pour certaines lignées, quelle que soit la température d'appariement employée. Il peut seulement être noté que les bandes sont plus nettes à 60°C. Pour les autres échantillons, seuls des smear sont visibles, pouvant être expliqués par un excès de matrice. Au final, seul 19d a des bandes qui apparaissent à la fois pour les côtés LB et RB. Pour 556-10 une bande se distingue pour le côté LB seulement, cependant beaucoup de bruit de fond est visible également.

Nous amplifions ensuite par PCR sur colonies seulement les fragments obtenus pour 19d LB et 19d RB, et obtenons des bandes de tailles attendues. Des bandes de tailles différentes sont visibles : pour 19d LB il y a des bandes de 900, 650, 400 et 350 pb ; pour 19d RB elles font toutes environ 450 pb (*voir Figure 12*). Vu que les amorces M13 sont utilisées, 200 pb doivent être enlevés pour correspondre à la taille de l'insert.

Pour chaque bordure, un site a pu être caractérisé. Par des recherches d'homologie de séquences sur Phytozome, la séquence d'une bordure gauche, mesurant 240 pb, a été trouvée comme étant positionnée au niveau du chromosome 5 (*voir Figure 13*). Par ailleurs, ce site d'insertion se trouve dans une région ne codant pour aucun gène. Concernant les autres amplicons séquencés, les recherches d'homologie de séquence n'ont pas abouties : il a été déterminé qu'elles avaient une forme complexe et comprenaient des régions du vecteur binaire pMDC32 répétées en tandem (résultats non montrés).

Parallèlement, une bordure droite est identifiée sur Phytozome comme étant sur le chromosome 14 (*voir Figure 14*). Ce site d'insertion se trouve cependant dans un gène : *Potri.014G041300*. Selon Phytozome, ce gène code pour une protéine « NAM » (No Apical Meristem), impliquée dans le développement des plantes.

Puits	1	2	3	4	5	6	7	8	9	10
Marqueur 1Kb	10d	15d	19d	23d	556- 10	10d	15d	19d	23d	556- 10
	LB					RB				

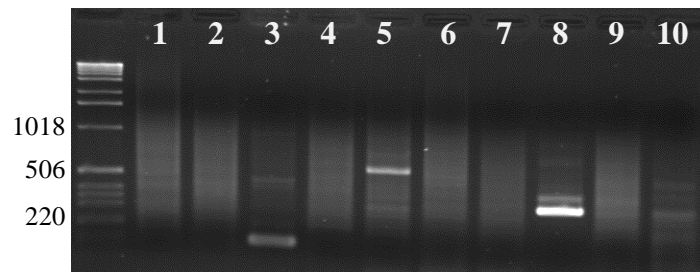


Figure 11 : Séparation sur gel d'agarose 1,5 % des amplicons obtenus par une 2^{ème} PCR « nested » de lignées avec pMDC32 en tant que vecteur binaire, digérées par HincII, avec une température d'appariement de 60°C. Dilution 2000x de la 1^{ère} PCR.

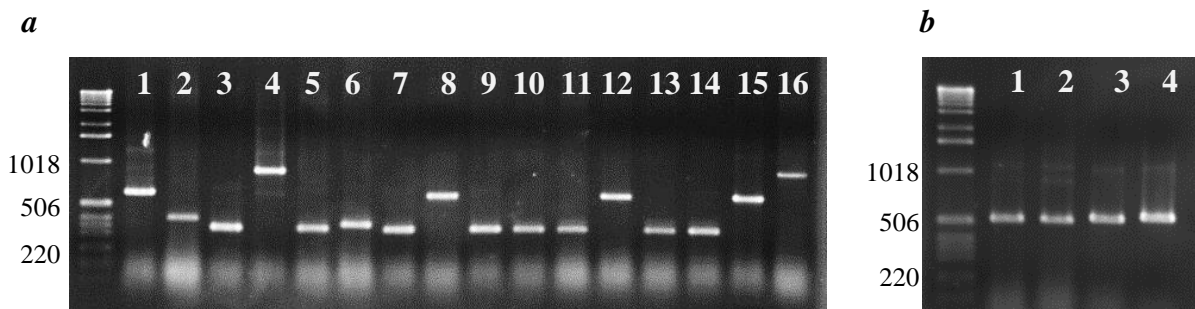


Figure 12 : Migration sur gels d'agarose 1,5% de PCR réalisées sur colonies pour (a) 19d LB et (b) 19d RB en utilisant les amorces M13

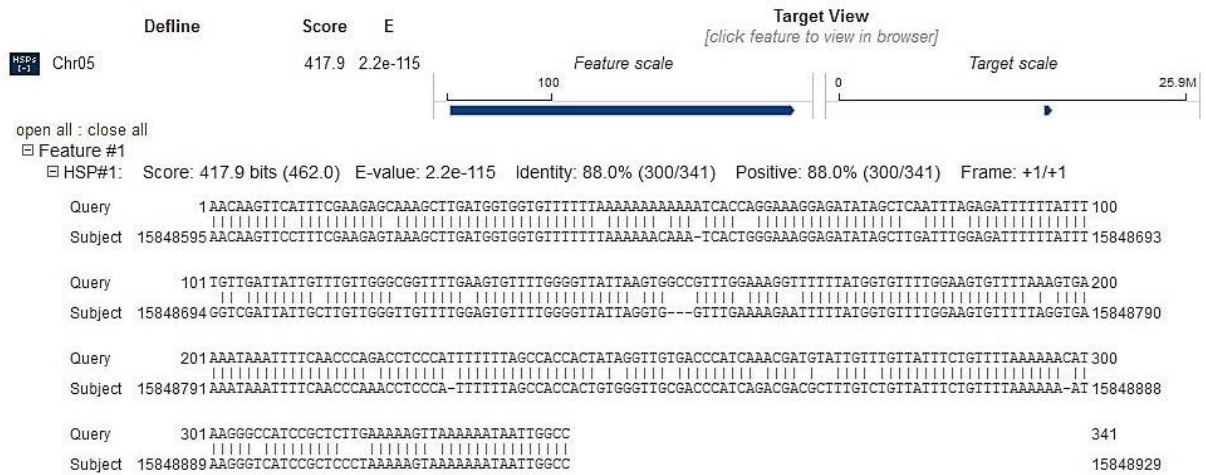


Figure 13 : Détermination d'un site d'insertion au niveau de la bordure gauche d'un ADN-T inséré chez 19d. Copie d'écran après analyse sur Phytozome montrant le résultat du BLAST. La bordure est positionnée sur le chromosome 5 de *Populus trichocarpa* (Phytozome)

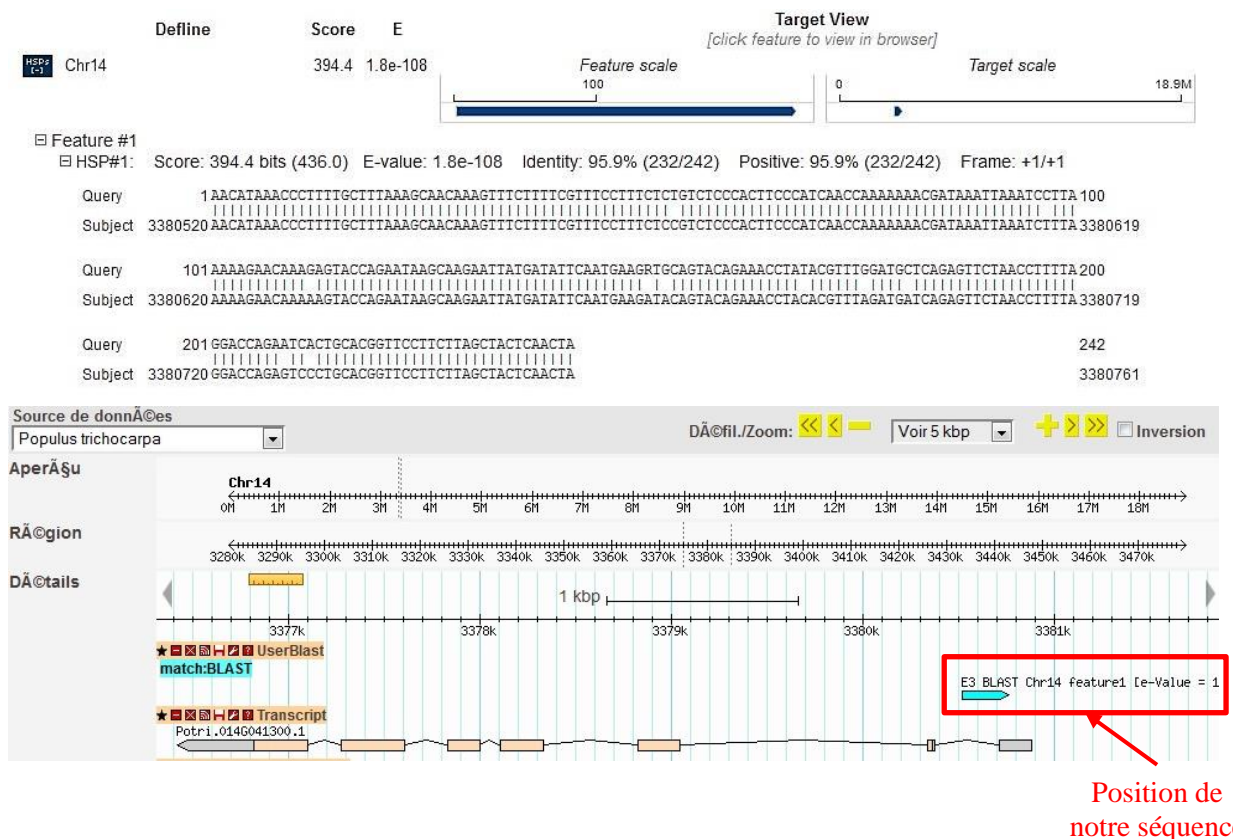


Figure 14 : Détermination d'un site d'insertion au niveau de la bordure droite d'un ADN-T inséré chez 19d. Copies d'écran après analyse sur Phytozome montrant le résultat du BLAST et sa position sur le chromosome 14 de *Populus trichocarpa* au niveau du gène *Potri.014G041300* (Phytozome)

2) Une sur-expression de miR397a impacte l'expression des laccases

Les différentes lignées sur lesquelles s'effectuent l'étude sont des lignées sur-exprimant miR397a, ainsi qu'une lignée témoin non transformée. Ces lignées sont analysées par RT-PCR quantitative, puis certaines d'entre elles sont sélectionnées pour être étudiées plus finement par RT-PCR quantitative. L'évaluation de l'expression des laccases s'est effectuée sur 4 gènes : *PtrLAC1*, *PtrLAC15*, *PtrLAC16* et *PtrLAC42*.

Les échantillons à analyser ont été amplifiés par RT-PCR, et ce avec les amorces PtrLac1, PtrLac16 et PtrLac15-42, ce dernier couple amplifiant à la fois *PtrLAC15* et *PtrLAC42*. D'autres amplifications ont été réalisées avec les amorces 18S pour le témoin. Pour *PtrLAC1*, nous pouvons voir des bandes d'intensités différentes suite à la migration des produits d'amplification sur gel (*voir Figure 15*). Par exemple, les lignées A1, A2, A3, A14 et A35 ont une intensité visiblement moins importante que pour le témoin non transformé. D'autres lignées comme A5, A7, A15, A23 et A25 ne montrent pas une différence d'expression par rapport au témoin. Lorsque les amorces PtrLac15-42 et PtrLac16 sont utilisées, les mêmes profils peuvent être observés. Ainsi, nous avons sélectionnés les lignées les plus représentatives pour l'ensemble des 4 gènes de laccases testés pour être analysées plus finement par RT-PCR quantitative (lignées A1, A2, A3, A5, A7, A14, A15, A23, A25 et A35).

Un gène de conjugaison à l'ubiquitine (Ubiquitin conjugating enzyme UBC10) est utilisé en tant que gène de ménage. Pour chacune des lignées, nous utilisons deux échantillons biologiques et un duplicat technique est effectué pour chaque échantillon. L'expression des différents gènes est mesurée, puis les résultats obtenus sont analysés : une moyenne des Ct est calculée pour chaque duplicat technique, pour ensuite obtenir le ΔCt ($\Delta Ct = Ct_{\text{transgénique}} - Ct_{\text{témoin}}$). Un ratio transgénique sur témoin est alors calculé par la formule $2^{-\Delta Ct}$. Ce ratio est ensuite normalisé avec le ratio obtenu d'après UBC10, en calculant le ratio de l'échantillon sur le ratio de l'ubiquitine. Les résultats pour les répétitions biologiques sont ensuite moyennés et représentés sous forme d'histogramme (*voir Table V* et *Figure 14*). A noter que, pour la lignée A3, il n'a pu y avoir de répétition biologique à cause de complications lors des manipulations.

D'après les résultats de RT-PCR quantitative, certaines lignées présentent une diminution au niveau de l'expression de toutes les laccases étudiées ici : les lignées A1, A3 et A7, qui ont un ratio inférieur à 0,5. Les lignées A5, A14 et A15 ont des ratios compris entre 0,5 et 1,5 pour toutes les laccases. Pour la lignée A2, les expressions des gènes des laccases *PtrLAC1* et

PtrLAC16 sont à un ratio de 0,6 environ ; cependant le ratio observé avec les amorces PtrLac15-42 est nettement supérieur et atteint 2,2. Les lignées A23, A25 et A35 sont celles où les ratios sont les plus importants : pour A35 par exemple, il atteint un ratio de 10,4 pour Lac15-42, et va même jusqu'à 25,2 pour Lac1. Le ratio observé pour Lac16 reste à 1,7. Pour A23 et A25, les ratios obtenus sont très similaires : environ 2,5 pour Lac1, 1,5 pour Lac15-42 et 2,5 pour Lac16 (voir Table V et Figure 16).

Entre les deux méthodes, RT-PCR semi-quantitative et RT-PCR quantitative, des inconsistances peuvent être notées. Pour la lignée A7 par exemple, il ne semblait pas y avoir de différence d'expression observée comparé à la lignée sauvage par RT-PCR semi-quantitative, mais lors de la RT-PCR quantitative les résultats montrent une diminution de l'expression des gènes codant pour toutes les laccases étudiées (avec des ratios entre 0,1 et 0,18). La lignée qui se démarque le plus est cependant A35 : elle a une amplification qui semble quasi inexistante sur gel, tandis que la quantification de son expression démontre le contraire, avec un ratio allant jusqu'à 25,2 pour *PtrLAC1*. Néanmoins, certaines observations concordent, notamment pour les lignées A1 et A3 où l'intensité des bandes semblaient faibles, et où la quantification le confirme avec des ratios allant de 0,19 à 0,48 pour A1 ; et de 0,3 à 0,4 pour A3.

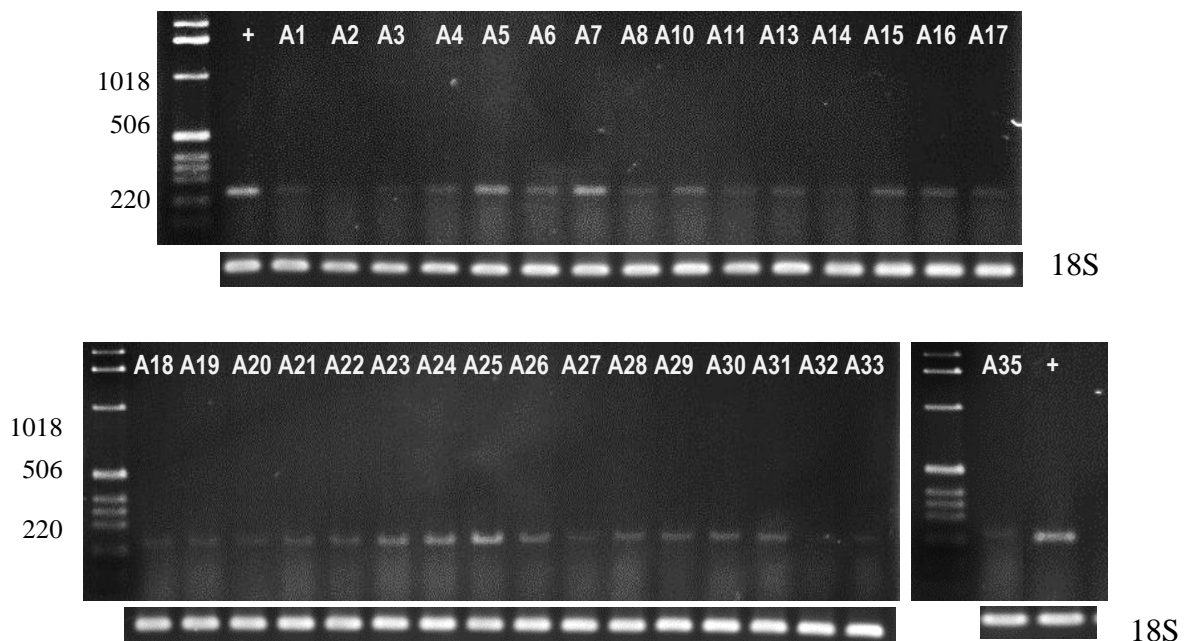


Figure 15 : RT-PCR semi-quantitative réalisée avec les amorces PtrLac1, et les amorces 18S pour le témoin, sur des échantillons de lignées surexprimant miR397a. (+) témoin non transformé

Table V : Ratios témoins sur transgénique moyennés entre les réplicas biologiques et normalisés, ainsi que les écart-types calculés entre les répétitions biologiques

	LAC1		LAC15		LAC16	
	moyenne	écart type	moyenne	écart type	moyenne	écart type
A1	0,478	0,145	0,186	0,229	0,322	0,199
A2	0,607	0,370	2,241	3,050	0,665	0,367
A3	0,381	-	0,297	-	0,336	-
A5	0,577	0,577	0,476	0,457	0,704	0,427
A7	0,104	0,034	0,176	0,240	0,165	0,059
A14	0,849	0,673	0,594	0,804	0,735	0,604
A15	1,212	0,047	0,948	0,471	0,988	0,115
A23	2,483	1,977	1,529	1,970	2,499	1,379
A25	2,551	2,235	1,674	1,180	2,577	1,807
A35	25,188	32,877	10,420	11,583	1,651	0,692

- Ratios inférieurs à 0,5
- Ratios entre 0,5 et 2
- Ratios supérieurs à 2

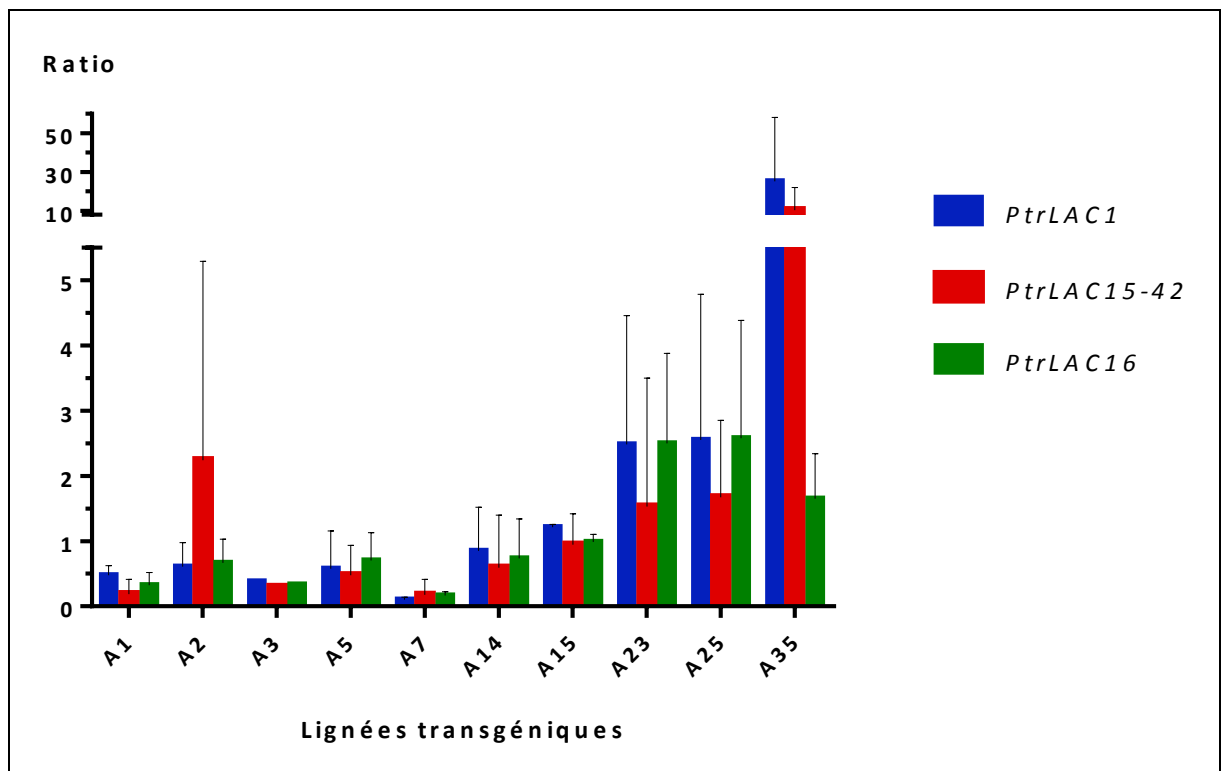


Figure 16 : Expression des gènes codant pour les laccases pour différentes lignées surexprimant miR397a. Les barres d'erreur représentent les écart-types observés entre les ratios des répétitions biologiques.

DISCUSSION

La caractérisation des sites d'insertion a pu aboutir à l'identification de la position de certains ADN-T insérés, en comparant notre séquence d'intérêt au génome de *Populus trichocarpa*.

Les premiers résultats obtenus pour FS40 et FS3 (*voir Figure 3*) montrent un profil similaire pour les 2 lignées en ce qui concerne la taille des fragments d'intérêt. L'hypothèse est une erreur d'étiquetage lors de la multiplication des lignées en culture *in vitro*. Des analyses de Southern refaites récemment avaient déjà révélée cette possibilité. Grâce aux résultats de séquençage, l'hypothèse est vérifiée : FS40 et FS3 sont une seule et même lignée. De plus, du fait que ces lignées sont connues du centre de recherche, il peut être affirmé que c'est la lignée FS3 qui a été perdue. Effectivement, les données de Southern montrent que le profil est le même que celui de la lignée FS40 originellement obtenue. Dans ce cas, la position des bordures gauche et droite de l'unique site d'insertion a pu être déterminée. Ainsi, les deux bordures se trouvent sur le même chromosome : le chromosome 6. Néanmoins, un éloignement important entre ces deux bordures est à constater : la bordure gauche se situe des positions 10825685 à 10825438, et la bordure droite de 12066670 à 12066763 sur le génome d'après Phytozome soit un éloignement des deux bordures de 1240985 pb, ce qui ne correspond pas à la taille de l'insert. D'après ces constatations nous pouvons émettre plusieurs hypothèses. Tout d'abord, une délétion peut être survenue dans le génome. Cependant, dans ce cas la délétion serait considérable : 1240985 pb, un peu moins si nous enlevons l'insert. Des délétions, voire même des insertions ont déjà été observées dans des études semblables (Kumar *et al*, 2002), toutefois, des délétions aussi importantes n'ont jamais été décrites à notre connaissance. Une autre hypothèse pouvant être posée serait la délétion d'une des deux bordures de l'ADN-T, qui se serait ensuite insérée plus loin dans le génome. Des problèmes de fonctionnement des bordures de l'ADN-T sont déjà connus (Ortega *et al*, 2002) et cette hypothèse semble solide. Enfin, il est également possible que le génome de *P. trichocarpa* n'ait pas été correctement cartographié, ou encore qu'une translocation/délétion ait eu lieu dans le génome de *Populus tremula x P. alba* par rapport à celui de *Populus trichocarpa*, ce qui pourrait expliquer l'apparent éloignement des deux bordures.

Pour la lignée FAS13, il a pu être observé qu'un site d'insertion se trouvait dans un gène codant pour une ubiquitine : il est donc possible que cette insertion affecte la physiologie de ces peupliers. Cependant, il faut également noter que cette ubiquitine fait partie d'une famille multigénique : si l'une d'entre elle est affectée, il est possible que sa fonction soit compensée par une autre ubiquitine. Pour FAS13 LB, les résultats laissaient supposer que plusieurs sites

d'insertion avaient été trouvés cependant les séquences retrouvées étaient les mêmes, à une exception près : certaines séquences étaient complexes et avaient, en plus de la séquence à identifier, une partie du vecteur pBIB HYGRO. Cette complexité pourrait provenir d'artefacts de clonage, mais elle pourrait également apparaître lors de la mobilisation et de l'insertion de l'ADN-T. Ces événements où l'on observe qu'une partie du vecteur s'insère avec l'ADN-T ont déjà été observés chez *Arabidopsis thaliana* (Martineau *et al*, 1994). Du côté de la bordure droite de l'ADN-T, un autre petit fragment a également pu être séquencé. Seulement, cette séquence pouvait se positionner dans tous les chromosomes, et ce à plusieurs endroits différents : il est possible que l'ADN-T se soit positionné au niveau d'une séquence répétée.

Enfin, pour la lignée Myb90-19d, deux sites d'insertion ont été trouvés, à chaque fois au niveau d'une seule bordure de l'ADN-T. Il est possible que des délétions en soient la cause, mais également que le site de restriction n'a pas été bien choisi. Dans tous les cas, l'un de ces sites d'insertion se situe au niveau d'un gène : *Potri.014G041300*. Il code pour une protéine NAM (No Apical Meristem) qui serait un régulateur transcriptionnel impliqué dans le contrôle des hormones. Cependant, dans les cas où l'on étudie des transformants primaires (hémizygotes au locus d'insertion), il reste peu probable de voir des effets sur le phénotype : il faudrait que les mutations soient dominantes, car il reste un allèle non muté par l'ADN-T. Lors de la caractérisation des sites d'insertion de la lignée 19d, il a pu être observé (comme pour la lignée FAS13) une complexité de séquences, avec une répétition en tandem d'une partie du vecteur binaire pMDC32.

Au final, la caractérisation des sites d'insertion n'est envisageable que pour des transgéniques ne présentant qu'un, ou tout au plus deux, site(s) d'insertion préalablement quantifié(s) en Southern ou par qPCR.

En parallèle, une étude a été menée sur des peupliers transgéniques sur-exprimant le micro-ARN miR397a, afin de voir s'il pouvait y avoir un impact sur l'expression des gènes des laccases *PtrLAC1*, *PtrLAC15*, *PtrLAC42* et *PtrLAC16*. Puisque miR397a est sur-exprimé et qu'il cible ces gènes codant des laccases, une sous-expression de ces derniers est attendue.

Par RT-PCR semi-quantitative des différences d'intensités sont visibles chez les plants transgéniques. D'après ces observations, dix lignées, A1, A2, A3, A5, A7, A14, A15, A23, A25 et A35, ont été sélectionnées pour être analysées par RT-PCR quantitative et vérifier qu'il y a bien des différences d'expression. Les résultats montrent, pour les lignées A1, A3 et A7, une sous-expression des gènes codant pour les laccases, avec des ratios inférieurs à 0,5. Ces résultats tendraient donc à montrer que miR397a régulerait les gènes codant les laccases

étudiées chez ces lignées. Toutefois, certains résultats de RT-PCR quantitative montrent un ratio d'expression supérieur à 2,5. L'expression des gènes chez ces lignées (par exemple, A23, A25 et A35) est plus importante que pour le témoin. Cependant, cette augmentation de l'expression peut être simplement due à la variabilité d'expression entre individus (variabilité biologique). Une analyse sur un plus grand nombre d'individus sera donc nécessaire pour ces lignées. Davantage d'individus sauvages devraient de même être analysés.

D'autres résultats sont surprenants, avec des incohérences entre les deux méthodes. Nous pouvons poser comme première hypothèse que ces différences sont dues à des erreurs de manipulations, comme une non reproductibilité au niveau des dépôts sur gel, pour l'analyse par RT-PCR semi-quantitative.

En conclusion, la méthode permettant de connaître la position exacte des sites d'insertion est difficile à mettre en œuvre lorsqu'il y a plusieurs sites d'insertion. Cette mise au point devrait toutefois permettre la caractérisation des peupliers transgéniques sur-exprimant miR397a. Ceci permettra de sélectionner parmi les lignées transgéniques celles ne présentant qu'un seul site d'insertion, si possible en dehors d'un gène, pour éviter, même si cela est rare, des effets de mutations insertionnelles.

BIBLIOGRAPHIE

- Abdel-Ghany S.E. and Pilon M.** (2008). MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. *J. Biol. Chem.* **283**: 15932–15945.
- Balergue S., Dubreucq B., Chauvin S., Le-Clainche I., Le Boulaire F., de Rose R., Samson F., Biaudet V., Lecharny A., Cruaud C., Weissenbach J., Caboche M., and Lepiniec L.** (2001). Improved PCR-walking for large-scale isolation of plant T-DNA borders. *BioTechniques* **30**: 496–498, 502, 504.
- Becker D.** (1990). Binary vectors which allow the exchange of plant selectable markers and reporter genes. *Nucleic Acids Res* **18**: 203.
- Berthet S., Demont-Caulet N., Pollet B., Bidzinski P., Cézard L., Le Bris P., Borrega N., Hervé J., Blondet E., Balergue S., Lapierre C., and Jouanin L.** (2011). Disruption of *LACCASE4* and *17* results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* **23**: 1124–1137.
- Berthet S., Thevenin J., Baratiny D., Demont-Caulet N., Debeaujon I., Bidzinski P., Leple J.C., Huis R., Hawkins S., Gomez L.-D., Lapierre C., and Jouanin L.** (2012). Role of Plant Laccases in Lignin Polymerization. *Advances in Botanical Research* **61** : 145–172.
- Bidlack J., Malone M., and Benson R.** (1992). Molecular Structure and component intergration of secondary cell walls in plants. *Proceedings of the Oklahoma Academy of Science.* **72**: 51–56.
- Boerjan W., Ralph J., and Baucher M.** (2003). Lignin Biosynthesis. *Annual Review of Plant Biology* **54**: 519–546.
- Christensen J.H., Overney S., Rohde A., Diaz W.A., Bauw G., Simon P., Montagu M.V., and Boerjan W.** (2001). The syringaldazine-oxidizing peroxidase PXP 3-4 from poplar xylem: cDNA isolation, characterization and expression. *Plant Mol Biol* **47**: 581–593.
- Curtis M.D. and Grossniklaus U.** (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**: 462–469.
- Déjardin A., Laurans F., Arnaud D., Breton C., Pilate G., and Leplé J.-C.** (2010). Wood formation in Angiosperms. *C. R. Biol.* **333**: 325–334.
- Devic M., Albert S., Delseny M., and Roscoe T.J.** (1997). Efficient PCR walking on plant genomic DNA. *Plant physiology and biochemistry* **35**: 331–339.
- Gelvin S.B.** (2003). Agrobacterium-Mediated Plant Transformation: the Biology behind the “Gene-Jockeying” Tool. *Microbiol. Mol. Biol. Rev.* **67**: 16–37.

- Goff S. A., Ricke D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., Sessions A., Oeller P., Varma H., Hadley D., Hutchison D., Martin C., Katagiri F., Lange B. M., Moughamer T., Xia Y., Budworth P., Zhong J., Miguel T., Paszkowski U., Zhang S., Colbert M., Sun W., Chen L., Cooper B., Park S., Wood T. C., Mao L., Quail P., Wing R., Dean R., Yu Y., Zharkikh A., Shen R., Sahasrabudhe S., Thomas A., Cannings R., Gutin A., Pruss D., Reid J., Tavtigian S., Mitchell J., Eldredge G., Scholl T., Miller R. M., Bhatnagar S., Adey N., Rubano T., Tusneem N., Robinson R., Feldhaus J., Macalma T., Oliphant A., Briggs S.** (2002). A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92–100.
- Goodstein D.M., Shu S., Howson R., Neupane R., Hayes R.D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., and Rokhsar D.S.** (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178–1186.
- Kumar S. and Fladung M.** (2002). Transgene integration in aspen: structures of integration sites and mechanism of T-DNA integration. *Plant J.* **31**: 543–551.
- Leple, J.-C., Dejardin, A., Laurans, F., Pilate, G., Goue, N., Label, P., Beritognolo, I., Boizot, N., and Breton, C.** *Physiologie et génomique de la formation du bois : Le peuplier à l'ère génomique.* *Biofutur* **247**: 43–48.
- Liu Y.G., Mitsukawa N., Oosumi T., and Whittier R.F.** (1995). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**: 457–463.
- Liu, Y.-G. and Whittier, R.F.** (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25**: 674–681.
- Martineau B., Voelker T.A., and Sanders R.A.** (1994). On Defining T-DNA. *Plant Cell* **6**: 1032–1033.
- Ortega D., Raynal M., Laudie M., Llauro C., Cooke R., Devic M., Genestier S., Picard G., Abad P., Contard P., Sarrobert C., Nussaume L., Bechtold N., Horlow C., Pelletier G., and Delseny M.** (2002). Flanking sequence tags in *Arabidopsis thaliana* T-DNA insertion lines: a pilot study. *Comptes Rendus Biologies* **325**: 773–780.
- Pilate G., Dejardin A., and Leple J.C.** (2012). Field trials with lignin-modified transgenic trees. *Advances in Botanical Research* **61**: 1–36.
- Plomion C., Leprovost G., and Stokes A.** (2001). Wood Formation in Trees. *Plant Physiol.* **127**: 1513–1523.
- Ranocha P., Chabannes M., Chamayou S., Danoun S., Jauneau A., Boudet A.-M., and Goffner D.** (2002). Laccase Down-Regulation Causes Alterations in Phenolic Metabolism and Cell Wall Structure in Poplar. *Plant Physiol.* **129**: 145–155.

- De la Riva G.A., Gonzalez-Cabrera J., Vazquez-Padron R. and Ayra-Pardo, C. (1998).** *Agrobacterium tumefaciens*: a natural tool for plant transformation. *Electronic Journal of Biotechnology* **1**: 118–113.
- Simmons B.A., Loqué D., and Ralph J. (2010).** Advances in modifying lignin for enhanced biofuel production. *Curr. Opin. Plant Biol.* **13**: 313–320.
- The Arabidopsis Genome Initiative (2000).** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Tuskan G.A., DiFazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A., Schein J., Sterck L., Aerts A., Bhalerao R.R., Bhalerao R.P., Blaudez D., Boerjan W., Brun A., Brunner A., Busov V., Campbell M., Carlson J., Chalot M., Chapman J., Chen G.-L., Cooper D., Coutinho P. M., Couturier J., Covert S., Cronk Q., Cunningham R., Davis J., Degroeve S., Dejardin A., Depamphilis C., Detter J., Dirks B., Dubchak I., Duplessis S., Ehlting J., Ellis B., Gendler K., Goodstein D., Gribskov M., Grimwood J., Groover A., Gunter L., Hamberger B., Heinze B., Helariutta Y., Henrissat B., Holligan D., Holt R., Huang W., Islam-Faridi N., Jones S., Jones-Rhoades M., Jorgensen R., Joshi C., Kangasjarvi J., Karlsson J., Kelleher C., Kirkpatrick R., Kirst M., Kohler A., Kalluri U., Larimer F., Leebens-Mack J., Leple J.-C., Locascio P., Lou Y., Lucas S., Martin F., Montanini B., Napoli C., Nelson D. R., Nelson C., Nieminen K., Nilsson O., Pereda V., Peter G., Philippe R., Pilate G., Poliakov A., Razumovskaya J., Richardson P., Rinaldi C., Ritland K., Rouze P., Ryaboy D., Schmutz J., Schrader J., Segerman B., Shin H., Siddiqui A., Sterky F., Terry A., Tsai C.-J., Uberbacher E., Unneberg P., Vahala J., Wall K., Wessler S., Yang G., Yin T., Douglas C., Marra M., Sandberg G., Van De Peer Y., Rokhsar D. (2006).** The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.
- Vanholme R., Demedts B., Morreel K., Ralph J., and Boerjan W. (2010).** Lignin Biosynthesis and Structure. *Plant Physiol.* **153**: 895–905.

SITOGRAFIE

blast.ncbi.nlm.nih.gov/

<http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html>

www.phytozome.net

www.invitrogen.com

<http://www.biology.utah.edu/> (pour le logiciel A Plasmid Editor)

<http://www.basic.northwestern.edu/biotools/OligoCalc.html>

ANNEXES

Organigramme de l'unité AGPF (au 15-05-2013)

Equipe de direction

G.Pilate (DU) & A. Déjardin (DUA)

Gestion : P. Montes (TRE)

Commandes & Communication : B. Viguier (TRS)

Ressources humaines : V Fréjaville (TRS)

Ressources Informatiques : JL Haton (AI)

Compétences techniques

Bioinformatique

O Rogier (IE2)

Chimie du bois

K Ader (TRN)

JP Charpentier (IR2, ACP)

Analyses protéiques

N Boizot (AI)

Analyses transcriptomiques

MC Lesage-Descauses
(AI)

Génotypage

V Guérin (AI)

Transgénèse & embryogenèse somatique

C Le Metté (TRN, ACP)

N Millet (TRE)

Histologie moléculaire & anatomie du bois

F Laurans (IR2, ACP)

V. Lainé-Prade (TRN)

Tests pathologiques

C. Ridel (TRN)

Compétences scientifiques

Physiologistes

C Breton (CR1)

A Déjardin (CR1, HDR)

M- Lelu-Walter (DR2, HDR)

JC. Leplé (CR1)

G Pilate (DR2, HDR)

C.Teyssier (CR1)

Post-docs & Doctorants

J Migeot (Post-doc)

G Dalla-Salda

D Gourcilleau (Post-doc)

F Guedes

J Guet

W Lakhhal

A Morel

M Nardin

N Richet (Post-doc)

M Ruiz Diaz Britez

C Wintenberger

Généticiens

C Bastien (DR2)

JC Bastien (IRC)

A Dowkiw (CR1)

J.Dufour (IR1 en mission)

V Jorge (CR1)

L Pâques (DR2, HDR)

P Rozenberg (DR2, HDR)

L Sanchez (CR1, HDR)

F Santi (CR1)

V Segura (CR2)

C Sindou (IR2)

M Villar (CR1, HDR)

CDD & stagiaires

M Bailly (CDD)

S Marin (CDD)

V Rong (CDD)

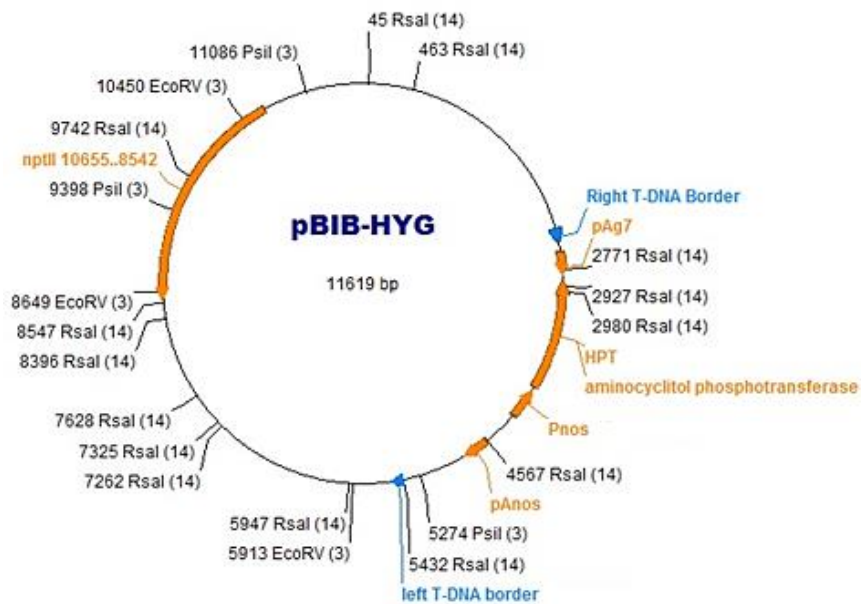
Responsables Activités transversales

Démarche Qualité : V Guérin ; **CHSCT** : JP Charpentier, N Boizot

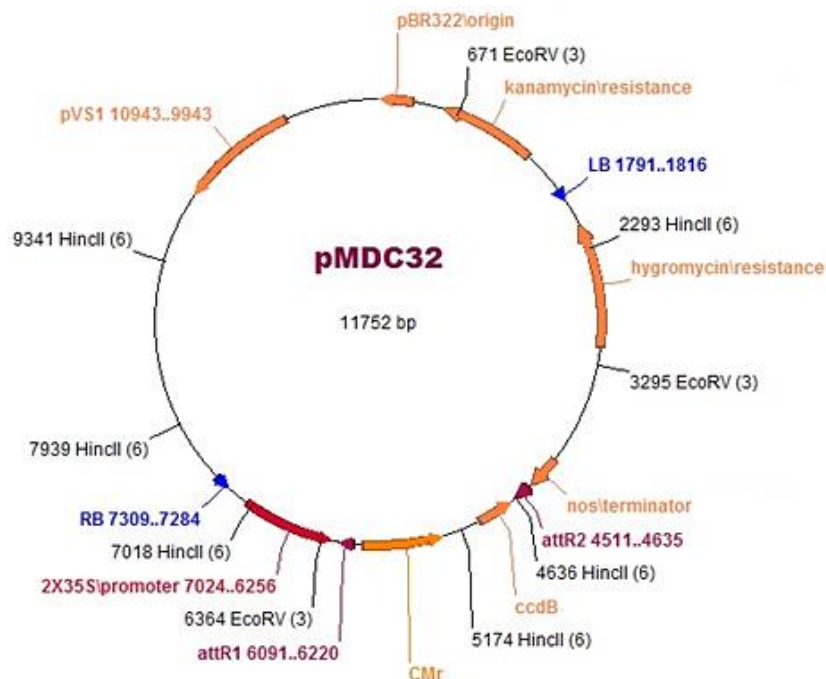
Corresp. formation : F Laurans

ANNEXE II

Cartes des plasmides pBIB HYGRO et pMDC32



A - Vecteur binaire pBIB HYGRO utilisé pour les plants FS40, FS3 et FAS13. Les sites de coupe des enzymes de restriction *RsaI*, *PstI* et *EcoRV* sont indiqués



B - Vecteur binaire pMDC32 utilisé pour les plants Myb156-10d, Myb156-15d, Myb90-19d, Myb90-23d et 556-10. Les sites de coupe des enzymes de restriction *EcoRV* et *HincII* sont indiqués.

ANNEXE III

Composition des tampons utilisés pour les migrations sur gels d'agarose

Tampon de charge 6X

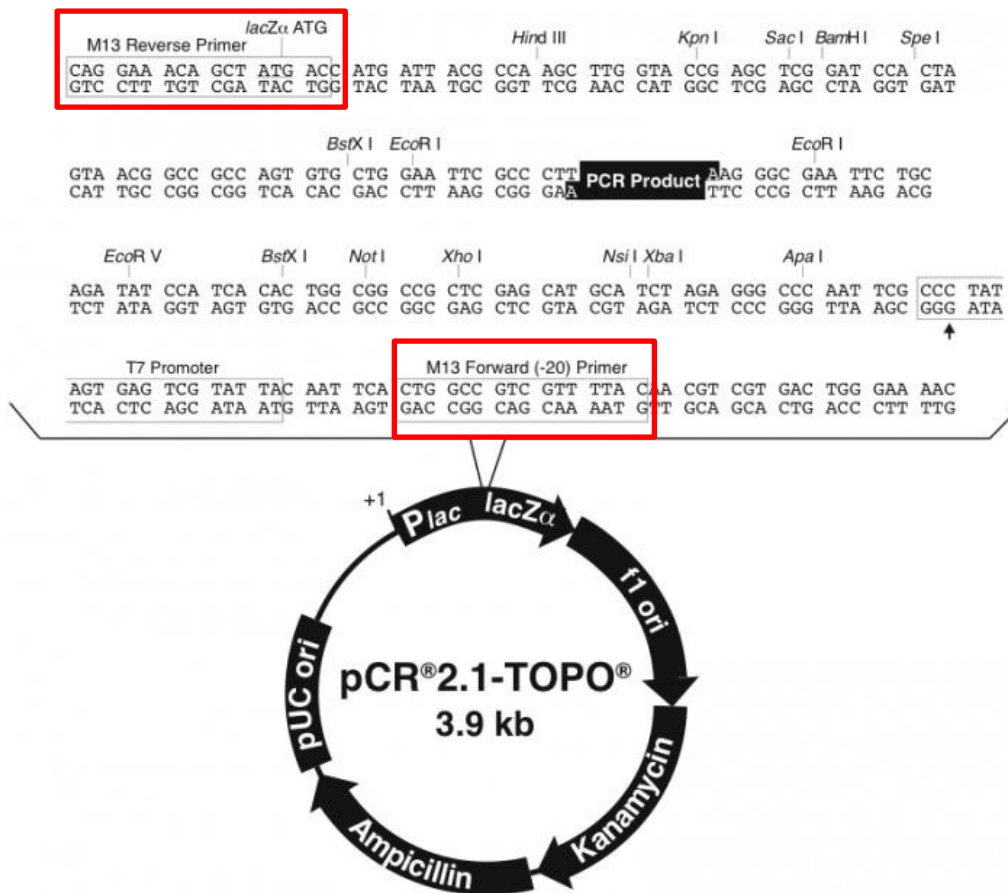
Bleu de bromophénol	0,25 %
Bleu de xylene cyanol	0,25 %
Glycerol	30 %

Tampon TBE10X (Tris, Borate EDTA)

Tris base	108 g
Acide borique	55 g
EDTA 0,5M pH8,0	40 mL
H2O	Qsp 1L

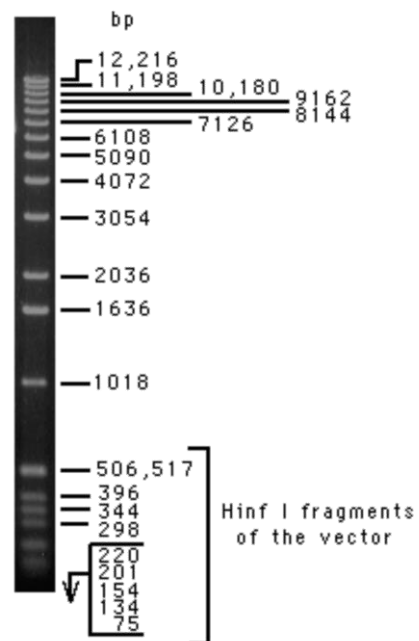
ANNEXE IV

Vecteur pCR®2.1-TOPO utilisé lors du clonage de produits PCR



ANNEXE V

Marqueur de taille 1kb ladder utilisé pour les gels d'agarose



RÉSUMÉ

L'intérêt général pour les bioénergies a conduit au développement de biocarburants de seconde génération *via* l'utilisation de biomasse lignocellulosique. Afin d'en apprendre davantage sur les gènes impliqués dans les propriétés du bois, des espèces à croissance rapide tel que le peuplier ont été étudiées et transformées génétiquement. Nous nous intéressons dans un premier temps à l'endroit où s'est inséré l'ADN-T d'*Agrobacterium tumefaciens* chez des peupliers transgéniques : c'est la caractérisation du (des) site(s) d'insertion. Cette méthode a d'abord été mise au point sur certaines lignées connues du laboratoire, comme FAS13. Nous avons pu constater, chez FAS13, qu'un ADN-T s'est inséré dans le chromosome 6 au niveau d'un gène d'ubiquitine *Potri.006G129600*. En parallèle, une analyse est réalisée sur des peupliers transgéniques sur-exprimant miR397a, afin de voir si ce micro-ARN a un impact sur l'expression de gènes codant des laccases. L'expression de 4 laccases représentatives, *PtrLAC1*, *PtrLAC15*, *PtrLAC16* et *PtrLAC42*, est suivie grâce à la RT-PCR semi-quantitative et la RT-PCR quantitative chez ces plants transgéniques. Trois de ces lignées sur-exprimant miR397a ont une diminution de l'expression des gènes codant pour les quatre laccases étudiées. Le but est désormais de caractériser les sites d'insertion de ces lignées.

Mots clés : Peuplier – micro-ARN - formation du bois - laccases - site d'insertion de transgène

ABSTRACT

The global interest in bio-energy has led to the development of second-generation biofuels through the use of lignocellulosic biomass. To learn more about the genes involved in wood properties, fast-growing tree species such as poplar were studied and genetically transformed. Firstly, we are interested in the place where the T-DNA of *Agrobacterium tumefaciens* is inserted in transgenic poplars: it's the characterization of the insertion site. This method was developed on some known strains of the laboratory, for example FAS13. We have seen, in FAS13, that a T-DNA is inserted into the 6th chromosome, in the ubiquitin gene *Potri.006G129600*. In parallel, an analysis is performed on transgenic poplars over-expressing miR397a, to observe any impact of this miRNA on the expression of genes encoding laccases. The expression of four representative laccases, *PtrLAC1*, *PtrLAC15*, *PtrLAC16* and *PtrLAC42*, are followed by semi-quantitative RT-PCR and quantitative RT-PCR in these transgenic plants. Three of these lines over-expressing miR397a have decreased expression of genes coding for the four laccases studied. The aim now is to characterize the insertion sites of these lines.

Keywords : Poplar - micro-RNA - wood formation – laccases - transgene insertion site