Architecture génétique des caractères cibles pour la culture du peuplier en taillis à courte rotation

Redouane El Malki

▶ **To cite this version:**

Redouane El Malki. Architecture génétique des caractères cibles pour la culture du peuplier en taillis à courte rotation. Sciences agricoles. Université d'Orléans, 2013. Français. NNT : 2013ORLE2005 . tel-00859626

HAL Id: tel-00859626

https://theses.hal.science/tel-00859626

Submitted on 9 Sep 2013

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

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

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE
SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT

Unité de recherche Amélioration Génétique et Physiologie Forestières

THÈSE présentée par :

Redouane EL MALKI

Soutenue le : 21 janvier 2013

pour obtenir le grade de : **Docteur de l'université d'Orléans**

Discipline/ Spécialité : Biologie

Architecture génétique des caractères cibles pour la culture du peuplier en taillis à courte rotation

THÈSE dirigée par :

Catherine BASTIEN	Directrice de Recherche, INRA d'Orléans

RAPPORTEURS :

Yves BARRIERE	Directeur de Recherche, INRA de Lusignan
Daniel PRAT	Professeur, Université Lyon I

JURY :

Stéphane MAURY	Professeur, Université d'Orléans – Président du jury
Maryse BRANCOURT	Chargée de Recherche, INRA de Mons
Jean-François RAMI	Chercheur, CIRAD de Montpellier
Yves BARRIERE	Directeur de Recherche, INRA de Lusignan
Daniel PRAT	Professeur, Université Lyon I
Catherine BASTIEN	Directrice de Recherche, INRA d'Orléans

“ *Dickmann and Keathley (1996), in questionning the role of physiology in breeding, stated that « … physiologists working with tree improvement programs for Populus and other taxa need to re-evaluate their modus operandi. » We strongly suggest that a new modus operandi might be found for physiological investigation in defining the relationship between trait expression and the economic worth therof* ”.
Riemenschneider et al, 2001

Remerciements

Tout d'abord, je tiens à remercier M. Gilles PILATE de m'avoir accueilli pendant ces trois ans au sein de l'unité Amélioration Génétique et Physiologie Forestière.
Je remercie aussi les membres du jury M. Yves BARRIERE, M. Daniel PRAT, M. Stéphane MAURY, Mme. Maryse BRANCOURT et M. Jean François RAMI d'avoir accepté de juger ce travail de thèse.
Je remercie le projet FUTUROL pour m'avoir financé cette thèse et pour l'organisation des échanges entre les différents partenaires du projet. Je remercie M. Jean Tayeb coordinateur du projet FUTUROL et M. Jean Charles Bastien (coordinateur du projet FURUROL au sein d'UAGPF), qui m'a accordé sa confiance pour cette thèse. Merci Jean Charles pour tes conseils et de ta bonne humeur (on a tué le virus)

Je tiens à remercier également les personnes qui ont participé aux deux comités de thèse organisés en fin de première et deuxième année et que leurs conseils et recommandations m'ont permis d'être efficace au cours de ces trois années. Pour le 1er comité de thèse, je remercie, Dominique Brunel, Camille Lepoittevin, Patricia Faivre Rampant, Jean Charles Bastien, Kevin Ader, Gilles Pilate, Anabelle Déjardin et Vanina Guérin
Pour le 2ème comité de thèse, je remercie Jean-Guy Berrin, Maryse Brancourt, Camille Lepoittevin, Patricia Faivre-Rampant, Gilles Pilate, Jean-Charles Leplé, Hélène Muranty, Annabelle Dejardin, Jean-Charles Bastien, Jean-Paul Charpentier, Vanina Guérin et Phillipe label.

Ce travail de thèse ne pourrait être accompli sans la contribution de trois personnes :
Tout ce que je vais dire ne saura exprimer ma gratitude et mon respect profond à vous trois :
Véronique, Catherine, Vincent, le trio VCV (Validation de la Cross Validation)
Ce trio a su être à la hauteur durant ces trois ans sur le côté scientifique, professionnel et humain.
Vous m'avez aidé jusqu'au bout pour surmonter les difficultés et aboutir à l'écriture de cette thèse (vous avez su assembler les QTL à effet faible des idées pour en faire qu'un QTL à effet fort d'idées rangées). Je vous remercie.
Le trio de choc :
Catherine, tu m'as appris la rigueur, malgré ton emploi de temps ministériel et tes obligations familiales (désolé Mathilde) tu étais là à m'écouter, à m'expliquer, à échanger et à m'aider à aboutir à ce travail. C'est le fruit de ton enseignement et de ta patiente.
Véronique, si je suis là en train d'écrire ces lignes, c'est grâce à toi. Tu m'as appris la progression. C'est toi qui m'a initié aux cartes génétiques et aux QTL, tu m'as appris plein de choses…
Vincent, tu m'as appris le challenge, ni « R » ni la génétique d'association n'ont pas de secrets à te cacher, merci pour tout, on aura le temps pour discuter encore et …

Cette thèse a été fertile à tous les niveaux, science et enfants. Bienvenue à Armel, Zoé et Firdaws que je souhaite qu'ils réussiront dans leurs vies futures et satisferont leurs parents

Un grand merci aux gens qui ont contribué au phénotypage et au génotypage :
Les composantes chimiques du bois :
Le plateau Génobois : Jean Paul Charpentier (merci pour tes explications et ta disponibilité), Nathalie Boizot (merci pour les raisins), Baptiste Favard, Flora Liégeois, Lucie Vernaux, Maxime Benardeau, Mathieu Migeon et Kevin Ader (merci pour le foot)
Le potentiel de saccharification
Je remercie également Jean Guy Berrin et David Navarro (les marseillais), les échanges avec vous étaient forts intéressants.

Phénotypage de la résistance
Patrick Poursat, Jean Gauvin, Virginie Dievart, José Almeida, Ludovic Pasquier, Daniel Lacan et Catherine Bastien.
Génotypage :
Vanina Guérin, Corinne Buret (j'ai failli tué ton poisson ! désolé!), Marie Claudes Lesage Descauses, Emanuelle Magnoux, Marine Herbert, Baptiste Favard et Flora Liégeois.

Je remercie mes collègues de l'unité qui ont rendu ce long séjour très agréable à leur côté.
Je commence par mes collègues de bureau avec qui j'ai partagé de bons moments, Isabelle bourgait (ton trèfle à quatre feuilles m'a porté chance), Nelly (merci pour les annotation v1.0, v2.0), Christelle Alluome, Johanne, Christelle Apha (merci beaucoup, Firdaws a failli naitre le même jour mais…), Elome, Sin yu (chenille), Marine (merci pour le livre que tu m'as offert)

Merci aux thésards : Julien (Biblio TCR, félicitation pour ton futur bébé !), Alexandre et Maxime (Tennis), Wassim (course) et Justine. Bon courage à vous tous pour la dernière ligne droite. Merci aux anciens thésards : Sylvain et Régis (Foot). Je souhaite aussi une bonne chance et une longue carrière pour les post-doctorants Jonathan, Nicolas et Delphine.

Merci aux collègues de table : Odile (Bienvenue à Orléans), Caroline (c'est fini les gels !), Philippe (le rythme des 100 m pour une distance du marathon vient de finir), Fernanda (j'ai découvert le Brésil sans y voyager, bon courage pour la suite), Lynda et Layla (coucou les algériennes)

Merci aux autres membres de l'unité : Patricia et Brigitte (efficacité administrative), Jean-Charles (pour les annotations), Gilles (conseils lors des comités de thèse et des présentations internes), Anabelle (une petite pensée à toi), Christian, Caroline, Philippe, Marie Anne (ta connaissance des oiseaux m'impressionne), Nathalie (merci pour les raisins), Frédérique (merci pour ton livre), Marc (Ton Saint Genix était superbe, merci pour le matériel et pour le partage de tes connaissances), Guillaume et Céline (pour la biblio TCR et TtCR), Franck et Karine (pour la bibliographie, leur bonne humeur, leur efficacité et leur disponibilité)

Je remercie également une personne qui a cru en moi et qui m'a donné envie de continuer dans la recherche
C'est Dominique Brunel, grâce à toi, je serais Docteur. Et je n'oublie pas de remercier une personne en particulier qui a suivi ce travail de thèse de près et qui a été derrière mon arrivée à Orléans, c'est Patricia Faivre R, merci.

J'ai surement raté quelqu'un, je m'excuse et je dis Merci à toute personne qui a contribué de proche ou de loin pour que ce travail se fasse en de très bonnes conditions

Merci à ma femme. Merci à toi Ikram de m'avoir supporté pendant ces trois ans, je suis enfin arrivé jusqu'au bout, grâce à ton aide et ton soutien. Maintenant, j'aurai plus de temps pour m'occuper de vous.

Merci à mes frères, à mon frère Mourad et à mes parents qui m'ont toujours poussé à faire cette thèse, je vous dois tout et je ne peux qu'exprimer mon humilité devant tout ce que vous avez sacrifié pour moi

Merci à sidi Mustapha Bassir et son épouse Mi Alhajja, ils seraient sans doute très contents s'ils étaient encore en vie !

LISTE DES PRINCIPALES ABRÉVIATIONS 1

LISTE DES TABLEAUX ET FIGURES 2

PRÉAMBULE 7

CHAPITRE I : COMMENT LA CRÉATION VARIÉTALE CHEZ LE PEUPLIER PEUT-ELLE RÉPONDRE À DES BESOINS ACCRUS EN RESSOURCE LIGNO-CELLULOSIQUE DÉDIÉE À LA PRODUCTION DE BIOCARBURANTS ? 13

A. VALORISATION DE LA RESSOURCE LIGNO-CELLULOSIQUE POUR LA PRODUCTION DE BIOCARBURANTS 13
1. Un contexte en faveur de la chimie verte et d'une utilisation accrue du bois comme source d'énergie 13
2. Quelle production de bois pour la production de biocarburants ? 19
B. POTENTIEL DES TAILLIS À COURTE ROTATION D'ARBRES 20
1. État de l'art, Agronomie et cultures dédiées (toutes espèces dont peuplier) 20
2. Exigences et limites actuelles des systèmes de culture en taillis pour la bioénergie 26
3. Nouveaux objectifs d'amélioration pour la culture en taillis de peuplier 28
C. RESSOURCES GÉNÉTIQUES DU PEUPLIER 38
1. Ressources cultivées 38
2. Exploration des ressources naturelles chez une espèce endémique *Populus nigra* (*P. nigra*) 42
D. UN PROGRAMME DE SÉLECTION EFFICACE À COURT TERME 47
1. Des contraintes à lever 47
2. Ressources génomiques disponibles chez le peuplier 48

CHAPITRE II : DÉVELOPPEMENT D'UNE BOITE À OUTIL MOLÉCULAIRE POUR LA CARTOGRAPHIE GÉNÉTIQUE CHEZ *P. NIGRA* 63

A. ÉTAT DE L'ART 63
1. Identification de polymorphismes SNP par reséquençage de gènes 63
2. Niveau et organisation de la diversité nucléotidique chez les peupliers 67
B. MATÉRIEL ET MÉTHODES 71
1. Matériel végétal 71
2. Choix des gènes candidats 71
3. Dessin des amorces 73
4. Conditions de la PCR et du séquençage 73
5. Analyse des séquences et de la diversité nucléotidique 74
6. Choix de SNP pour la construction d'un outil de génotypage SNPLex® destiné au pédigrée de cartographie génétique 76
C. RÉSULTATS 76
1. Le choix initial et le nombre final de fragments de gènes candidats analysés 76
2. Diversité nucléotidique 80
3. Déséquilibre de liaison 84
4. Liaisons entre indices de diversité et déséquilibre de liaison 84
5. Relations entre indices de diversité et pourcentage d'introns des fragments 84
6. Choix de SNP pour la construction d'un outil de génotypage SNPLex® destiné au pédigrée de cartographie génétique 89
D. DISCUSSION 89
1. Une stratégie de découverte de SNP validée 89
2. Gènes candidats ou sans *a priori* ? 91
3. Des niveaux de diversité et de déséquilibre de liaison très variables 92
E. PERSPECTIVES 92

CHAPITRE III : CARTES GÉNÉTIQUES CHEZ *P. NIGRA* 95

A. ÉTAT DE L'ART 95

1. Des marqueurs et des cartes génétiques 95
2. Cartes génétiques développées chez les Salicacées 103
3. Intérêt de la construction de nouvelles cartes génétiques chez *P. nigra* 109
B. Outils de conception de cartes génétiques utilisés 109
1. Famille de cartographie 109
2. Optimisation du choix des marqueurs 109
3. Élaboration des cartes génétiques 113
C. Cartes génétiques d'un pédigrée F1 de peuplier noir (Résultats) 115
1. Criblage de marqueurs et ségrégation 115
2. Construction des cartes génétiques parentales et consensus 119
3. Analyse par enrichissement « Gene Ontology » 123
D. Intérêt de ces nouvelles cartes génétiques (Discussion) 123
1. Ancrage de la carte génétique du *P. nigra* sur la carte physique de *P. trichocarpa* et intérêt pour l'assemblage *de novo* de *P. nigra* 123
2. Génétique comparée peuplier/saule 125
3. Distorsion 125
4. Conclusion : intérêt des cartes pour la cartographie de QTL 126

CHAPITRE IV : CARTOGRAPHIE DE QTL ASSOCIÉS À LA RÉSISTANCE PARTIELLE À LA ROUILLE FOLIAIRE 128

A. Introduction 131
B. Material and Methods 133
1. Plant material 133
2. Rust resistance assessments 133
3. Genotyping 134
4. Genetic map construction 136
5. Resistance data analysis 137
6. QTL detection 138
C. Results 140
7. Marker screening and genetic map construction 140
8. Phenotypic variability for rust resistance epidemiological components 143
9. QTL detection 150
D. Discussion 152
1. Added value of new *P. nigra* genetic maps 152
2. First identification of QTL associated to Mlp rust resistance in *P. nigra* 154
3. A strong genotype by strain interaction 155
4. A combination of major and minor resistance loci revealed in *P.nigra* 156
5. Phenotypic dissection of partial resistance to *Mlp* in *P. nigra* 157
E. Conclusion 158

CHAPITRE V: CARTOGRAPHIE DE QTL ASSOCIÉS À LA QUALITÉ DE LA BIOMASSE 161

A. Méthodes d'évaluation directes et indirectes des composants chimiques du bois 161
B. Analyse du déterminisme génétique et moléculaire de la qualité de la biomasse : état de l'art chez le peuplier 163
C. Matériel et méthodes 167
1. Préparation des échantillons 167
2. Acquisition des spectres et prétraitements statistiques 167
3. Analyses chimiques et potentiel de saccharification 169
4. Modèles de calibration 172
5. Set de validation 173
6. Héritabilités et corrélation entre caractères 173
7. Détection de QTL 175
D. Résultats 177

1. Analyse exploratoire de la variabilité phénotypique dans les échantillons de référence 177
2. Calibrations 181
3. Phénotypes prédits, validations, héritabilités et corrélation entre caractères prédits 181
4. Détection des QTL 191
E. Discussion 193
1. Intérêt de la SPIR comme méthode de phénotypage indirecte des propriétés chimiques et technologiques du bois 195
2. Influence de la composition chimique du bois sur le rendement en saccharification 196
3. Le déterminisme génétique des propriétés chimiques et technologiques du bois est complexe 197
4. Perspectives 198

CHAPITRE VI : DISCUSSION GÉNÉRALE ET PERSPECTIVES 201

A. Méthodes de découvertes de polymorphisme 201
B. Analyse de ce polymorphisme nucléotidique 202
1. Intérêts et perspectives de l'analyse de la seule diversité moléculaire 202
2. Un phénotypage de qualité qui reste encore coûteux 204
3. Vers la génétique d'association ? 205
C. Deux caractères ? Deux défis ? Deux déterminismes différents ? 210

ANNEXE 213

BIBLIOGRAPHIE 221

Liste des principales abréviations

ACP	**A**nalyse en **C**omposantes **P**rincipales
Ademe	**A**gence **D**e l'**E**nvironnement et de la **M**aîtrise de l'**É**nergie
ADN	**A**cide **D**ésoxyribo-**N**ucléique
ADNc	**A**cide **D**ésoxyribo-**N**ucléique **C**omplémentaire
AFLP	**A**mplification **F**ragment **L**ength **P**olymorhism
ARNm	**A**cide **R**iboNucléique **m**essager
CNV	**C**opy **N**umber **V**ariation
CIP	**C**ommission **I**nternationale du **P**euplier
EST	**E**xpressed **S**equence **T**ag
FCBA	Institut technologique de **F**orêt, **C**ellulose, **B**ois-construction et **A**meublement (AFOCEL anciennement)
GBS	**G**enotyping **B**y **S**equencing
IEA	**I**nternational **E**nergy **A**gency
IFN	**I**nstitut **F**orestier **N**ational
IFPEN	**I**nstitut **F**rançais du Pétrole **É**nergies **N**ouvelles
Indel	**In**sertion/**dél**étion
IPGC	**I**nternational **P**opulus **G**enome **C**onsortium
JGI	**J**oint **G**enome **I**nstitute
LOD	**L**ogarithm of the **od**ds
MAI	**M**ean **A**nnual **I**ncrement ou accroissement annuel moyen
miARN	**mi**cro **A**cide **R**iboNucléique
MS	**M**atière **S**èche
MWh	**M**éga**W**att**h**eure
NB-LRR	**N**ucleotide **B**inding **L**eucine-**R**ich **R**epeat
NGS	**N**ouvelle **G**énération de **S**équençage
ORNL	**O**ak **R**idge **N**ational **L**aboratory
QTL	**Q**uantitative **T**rait **L**oci
SAM	**S**élection **A**ssistée par **M**arqueurs
SNP	**S**ingle **N**ucleotide **P**olymorphism
SSR	**S**hort **S**equence **R**epeat ou microsatellites
TCR	**T**aillis à **C**ourte **R**otation
tep	**t**onne **é**quivalent **p**étrole
TtCR	**T**aillis à **t**rès **C**ourte **R**otation

Liste des tableaux et figures

Figure 0.1 : Représentation schématique des ressources génétiques de peuplier noir disponibles et utilisées dans la thèse 9

Figure I.1 : Les filières de valorisation énergétique de la biomasse en biocarburants..... 15

Figure I.2 : Les différents modules du projet Futurol avec les actions à accomplir 18

Tableau I.1 Productivité moyenne en taillis à courte rotation (TCR) et très courte rotation (TtCR) de 5 espèces forestières majoritairement utilisées en zones tempérées..... 23

Tableau I.2 : Évaluation des obstacles techniques et non techniques des taillis à courte rotation dédiés pour la bioénergie..... 27

Figure I.3 : Présentation schématique des principales composantes morphologiques et physiologiques affectant la croissance du peuplier conduit en taillis à courte rotation 32

Figure I.4 : Représentation des 3 molécules d'hydroxy-cinnamyl alcools à la base de la structure moléculaire des polymères de lignine. 35

Figure I.5 : Subdivision du genre *Populus* en plusieurs sections. *P. deltoides*, *P. trichocarpa* et *P. nigra* sont les 3 principales espèces utilisées en amélioration génétique...……..…37

Figure I.6 : Compatibilité de croisements entre espèces et sections du genre *Populus* 39

Figure I.7 : Caractéristiques botaniques du peuplier noir (Köhler's Medizinal-Pflanzen 1897)... 41

Figure I.8 : Aire de répartition naturelle du peuplier noir 41

Tableau I.3 : Niveaux de variabilité génétique et d'héritabilités estimés chez *P. nigra* pour différents caractères d'intérêt 44

Figure I.9 : La version 2.2 de l'assemblage du génome du peuplier..... 50

Tableau I.4 : Liste des études expressionnelles de gènes candidats associés aux mécanismes de défense conduites chez le peuplier 53

Figure I.10 : Représentation schématique des principales régulations du transcriptome du peuplier et de *Melampsora larici-populina* dans le cadre de l'interaction compatible..... 55

Figure I.11 : Représentation schématique sur les 19 chromosomes du peuplier des QTL (boites vertes et blanches), des gènes candidats fonctionnels (en rouge) et des gènes candidats expressionnels (points verts) associés à l'arrêt de croissance chez le peuplier..... 56

Tableau I.5 : Ressources génomiques du peuplier 58

Tableau II.1 : Patrons de diversité nucléotidique chez les salicacées 65

Tableau II.2 : Nombre de gènes communs entre les études de diversité dans le genre *Populus* citées dans le tableau II.1. 68

Figure II.1 : Origines géographiques des 21 individus *Populus nigra* utilisés pour le re-séquençage de gènes.....70

Figure II.2 : Schéma de la stratégie utilisée pour l'identification de SNP chez *Populus nigra*…….....72

Figure II.3 : Répartition des gènes candidats choisis a) sur tout le génome, et b) par catégorie ..72

Figure II.4 : Position physique des 665 fragments amplifiés sur la version 2.0 du génome de *P. trichocarpa* 75

Tableau II.3 : Nombre de gamètes, nombre d'haplotypes et diversité nucléotidique moyens (π) estimés à partir de 456 fragments de gènes dans un panel de 21 *P. nigra* 77

Figure II.5 : Distributions et boîtes à pattes : a) de la diversité nucléotidique et b) du nombre d'haplotypes pour 456 fragments de gènes (34,9 gamètes en moyenne) re-séquencés chez *P. nigra*, et c) du déséquilibre de liaison (r^2) moyen pour 369 fragments de gènes.....78

Tableau II.4 : Les gènes candidats à valeurs extrêmes extraits de l'analyse de la diversité dans chaque classe (π > 0,0074), avec leur annotation v2.0 du peuplier et l'annotation correspondante d'*A. thaliana* (tair).....79

Tableau II.5 : Les gènes candidats extrêmes extraits de l'analyse du nombre d'haplotypes dans chaque classe (Nb d'haplotypes >10), avec leur annotation v2.0 du peuplier et l'annotation correspondante d'*A. thaliana* (tair). 81

Figure II.6 : Boîte à pattes des valeurs de diversité nucléotidique pour chaque catégorie de gènes 82

Figure II.7 : Relation entre les valeurs de déséquilibre de liaison r^2 et la distance en paire de bases (369 fragments, nombre de points : 11716).....83

Figure II.8 : Distribution du déséquilibre de liaison (r^2) moyen par fragment pour les gènes candidats et pour les gènes choisis sans a priori (random). 85

Figure II.9 : Relation entre la diversité nucléotidique, le nombre d'haplotypes et le déséquilibre de liaison (r^2) moyen 86

Figure II.10 : Relation entre les 3 indices de diversité et le pourcentage d'intron dans les fragments de gènes : a) diversité nucléotidique ; b) nombre d'haplotypes ; c) déséquilibre de liaison (r^2) moyen.....87

Tableau II.6 : Potentiel de réussite du génotypage de 96 marqueurs SNP (SNPLex) identifiés à partir des deux parents 71072-501 et BDG dans les familles du factoriel de croisement 4x5 *P. nigra* 88

Tableau II.7 : Gènes communs à la présente étude et aux études sur la diversité de gènes candidats chez le genre *Populus*.....90

Tableau III.1 : Liste des différents logiciels de cartographie génétique avec leur disponibilité, statut et type de pédigrées traités.....99

Figure III.1 : Représentation schématique des populations de cartographie génétique 102

Tableau III.2 : Synthèse des cartes génétiques chez les peupliers 104

Figure III.2 : Les notes d'infection à la rouille foliaire observées en 2000 pour chaque famille d'un plan factoriel *P. nigra* 5x4 108

Figure III.3 : Représentation schématique des cas de figure possibles de SNP 111

Figure III.4 : Représentation schématique du codage double inversé des marqueurs A et B 114

Tableau III.3 : Les différents types de marqueurs polymorphes chez le parent femelle et le parent mâle du pédigrée F1 *P. nigra* 116

Tableau III.4 Bilan des caractéristiques des trois cartes génétiques (femelle, mâle et consensus) établies sur un pédigrée F1 *P. nigra* 116

Figure III.5 : Alignement sur le génome de *P. trichocarpa* Nisqually-1 v2.0 des cartes génétiques parentales basées sur l'analyse de la ségrégation de marqueurs SSR, SNP et AFLP dans la descendance du croisement entre *P. nigra* 71072-501 x BDG (319 individus F1). 117

Figure III.5: (suite) 118

Tableau III.5 : Discordances entre les cartes établies dans le pédigrée F1 *P. nigra*, les cartes génétiques publiées et construites sur la base de croisements impliquant au moins un parent *P. nigra* (Gaudet et al. 2008 et Cervera et al. 2001), et les différentes versions d'assemblage du génome *P. trichocarpa* (Tuskan et al. 2006 ; www.phytozome.net). 120

Tableau III.6 : Liste des 20 premiers termes GO (Gene Ontology) pour lesquels les chromosomes 6 et 18 du peuplier sont significativement enrichis 122

Table IV.1 : Summary of number of AFLP, SSR and SNP markers tested and genotyped in the F1 *P. nigra* mapping pedigree. 139

Table IV.2 : Summary of genetic maps characteristics. 139

Table IV.3 : Marker positions with discordant localization (non syntenic) between present work, genetic maps of (Gaudet et al. 2008) and (Cervera et al. 2001) and version 1.0, 2.0 and 3.0 of Nisqually-1 genome assembly. 141

Figure IV.1 : Distributions of genotypic means for latent period (LP, days), uredinia number (UN) and uredinia size (US, 1-5 scale) following artificial inoculation of the F1 progeny with the two *Melampsora larici-populina* strains 93JE3 and 98AG69 145

Table IV.4 : Genetic variances with standard deviation (SD), broad sense heritabilities (H^2) with standard error (SE) and relative contributions of genetic, genotype x strain (GS) interaction and residual variances in phenotypic variation in the F1 progeny for latent period (LP), uredinia number (UN) and size of uredinia (US) 147

Table IV.S2 : Parental and F1 progeny means with standard error (SE) for latent period (LP, days), uredinia number (UN) and size of uredinia (US, 1-5 scale) 147

Figure IV.2 : Relationships between genotypic means for latent period (LP, days), uredinia number (UN) and size of uredinia (US, 1-5 scale) following artificial inoculation of the F1 progeny with the two *Melampsora larici-populina* strains 93JE3 and 98AG69 148

Table IV.5 : QTL detected on the consensus framework map using MQM method from Joinmap and significant markers using a linear model associated with the variation of three epidemiological traits 149

Figure IV.3 : Position of QTL controlling epidemiological components of resistance to two *Melampsora larici-populina* strains 151

Figure IV.S2 : Boxplot of genotypic means for each allelic classes for all significant trait x marker association detected in the F1 mapping progeny inoculated with each of the two Mlp strains, 98AG69 and 93JE3. a. Latent period. b. Uredinia number (UN). c. Uredinia size (US). 153

Figure V.1 : Principe de la spectroscopie en proche et moyen infrarouge..... 160

Tableau V.1 : Gamme de variation de la teneur en lignines, du ratio S/G, des teneurs en sucres à cinq (C5) et six atomes de carbones (C6) avec leurs valeurs d'héritabilités au sens large (H^2) chez différentes espèces et hybrides de peuplier..... 162

Tableau V.2 : Synthèse bibliographique des associations significatives de SNP appartenant à des gènes candidats avec la teneur en lignines, le ratio S/G, la teneur en sucres à six carbones (C6) chez *P. trichocarpa* (Wegrzyn et al. 2010) et chez *P. nigra* (Guerra et al. 2013) 165

Figure V.2 : Représentation schématique des analyses (dosages et spectrométrie) effectuées pour phénotyper les caractères de qualité du bois des 479 échantillons du pédigrée F1 *P. nigra*. 168

Figure V.3 : Représentation schématique du protocole de préparation du résidu pariétal permettant le dosage de la teneur en extractibles 170

Figure V.4 : Représentation schématique du protocole de dosage de la teneur en lignines 170

Tableau V.3 : Analyse exploratoire de la variabilité phénotypique des caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra*. 174

Figure V.5 : Analyse en composantes principales des caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra* 176

Tableau V.4 : Modèles de calibration sélectionnés pour les caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra*. 178

Figure V.6 : Représentation graphique des validations croisées de Monte-Carlo pour les modèles de calibrations sélectionnés pour les caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra*. 179

Figure V.7 : Illustration de l'effet « date » sur les spectres statistiquement pré-traités (normalisation et dérivée 1ère) acquis sur les 479 échantillons du pédigrée F1 *P. nigra*..... 180

Figure V.8 : Représentation graphique de l'ajustement des valeurs prédites à l'effet « date » pour quelques caractères..... 182

Tableau V.5 : Coefficients de détermination des régressions entre les valeurs mesurées et prédites (avec ou sans ajustement à l'effet « date » de prise des spectres) sur les 16 échantillons de validation du pédigrée F1 *P. nigra* 183

Figure V.9 : Représentation graphique des validations des modèles de calibrations sur 16 échantillons de validation du pédigrée F1 *P. nigra* 184

Tableau V.6 : Variabilité phénotypique et héritabilités au sens large de différentes propriétés chimiques du bois dans un pédigrée F1 de *P. nigra* 186

Tableau V.7 : Matrice de corrélations phénotypiques entre caractères mesurés sur les 100 échantillons de référence (en-dessous de la diagonale) et entre les moyennes clonales des caractères prédits (au-dessus de la diagonale) 187

Figure V.10 : Analyse en composantes principales des moyennes clonales des caractères prédits sur les 479 échantillons du pédigrée F1 *P. nigra* 188

Tableau V.8 : Positions et statistiques associées aux QTL détectés en MQM (Multiple QTL Mapping) avec MapQTL sur les cartes jalon femelle, mâle et consensus d'un pédigrée F1 *P. nigra*. 189

Figure V.11 : Représentation graphique des QTL cartographiés avec MapQTL 190

Figure V.12 : Représentation graphique des effets opposés des 2 QTL associés à la teneur en lignines et à la quantité de sucres totaux 192

Tableau V.9 : Ensemble des QTLs associés aux caractères de propriétés du bois détectés avec MapQTL et R dans un pédigrée F1 *P. nigra* pour leurs effets propres ou épistatiques 194

Figure VI. 1 : Représentation graphique de l'analyse en composantes principales effectuée sur les moyennes clonales des caractères de la résistance à la rouille foliaire et la qualité du bois . 209

Tableau VI.1 : Performances moyennes pour des caractères de résistance à la rouille foliaire et de propriétés du bois de l'ensemble du pedigree et d'un sous-ensemble de génotypes sélectionnés 211

Préambule

Le projet FUTUROL dans lequel s'inscrit cette thèse vise le développement et la commercialisation d'un procédé complet de production de bioéthanol de 2ème génération à partir de biomasse ligno-cellulosique issue d'une grande variété de matières premières. Il est organisé en six axes de recherche prioritaires (ou modules) s'intéressant d'amont en aval à: (1) la caractérisation des différentes ressources végétales disponibles et à l'analyse de leurs systèmes de production, (2) à (5) aux différentes étapes industrielles de transformation de ces ressources en bioéthanol (prétraitement, hydrolyse avec sa mise au point de production d'enzymes à haut rendement, fermentation) et (6) à l'intégration de toutes ces étapes et à l'analyse du cycle de vie de ce nouveau processus industriel. Ce travail de thèse vise à faciliter le développement à court terme de nouvelles variétés clonales de peuplier adaptées à une utilisation en taillis à courte rotation (TCR) ou à très courte rotation (TtCR) et permettant la production d'une ressource de qualité en s'intéressant plus particulièrement à l'architecture génétique de caractères cibles en sélection et à l'élaboration d'outils moléculaires permettant une sélection à la fois plus rapide et plus efficace. Il s'inscrit donc dans la partie amont du projet dans le module s'intéressant aux ressources ligno-cellulosiques et a été conduit au sein de l'Unité de recherches INRA 588 « Amélioration, Génétique et Physiologie forestières » à Orléans depuis le 1er Novembre 2009.

Les variétés clonales de peuplier actuellement disponibles sur le marché ont été développées pour des usages différents. Afin d'optimiser les rendements des futurs TCR, il convient de faire porter les efforts de sélection sur les critères spécifiques à ces types de culture à savoir : l'aptitude au rejet de souche, la productivité juvénile, la répartition de la biomasse dans la cépée, la tolérance à une forte compétition, la résistance aux maladies dont le développement est favorisé par une densité élevée de plantation et bien-sûr les composantes qualitatives de cette biomasse.

L'obtention, à court terme, de gains génétiques ciblés sur ces nouveaux caractères chez le peuplier nécessite le développement de nouvelles méthodes de sélection plus efficaces. Seules des méthodes de sélection indirectes peuvent permettre de réduire temps et coût d'évaluation de caractères technologiques nouveaux non encore stabilisés et d'associer un grand nombre de critères différents. Ces dernières doivent concilier précision d'évaluation élevée pour des caractères complexes dont l'héritabilité et le niveau de variabilité génétique sont jusqu'ici peu connus et haut-débit pour permettre de valoriser une forte intensité de sélection en permettant le tri rapide d'un très grand nombre de candidats. Dans ce contexte, les marqueurs moléculaires se

révèlent des outils de choix s'ils sont bien choisis pour un niveau de déséquilibre de liaison suffisant avec les gènes contrôlant les caractères d'intérêt et pour leur niveau de polymorphisme (Lande and Thompson 1990; Meuwissen et al. 2001). En profitant des ressources génomiques disponibles chez le peuplier, cette thèse vise également à développer une boîte à outils moléculaires permettant de balayer l'ensemble du génome de peuplier noir (*P. nigra* L.) afin d'identifier les régions génomiques, les gènes candidats et les polymorphismes impliqués dans ces caractères d'intérêt. En termes de matériel végétal, elle repose d'une part sur l'étude de la diversité moléculaire au sein d'un panel de 21 individus *P. nigra* échantillonnés dans des populations naturelles françaises et d'autre part sur la cartographie de QTL dans un pédigrée de référence constitué de 324 individus plein-frères clonés issus du croisement entre deux parents inclus dans le panel de séquençage (**Figure 0.1**). Il est à noter que ce matériel végétal est connecté d'une part à une population d'association composée de 1100 individus couvrant une diversité géographique plus large et d'autre part à un plan de croisement factoriel cloné 5x4. Les résultats acquis dans cette thèse devraient ainsi permettre de préciser les modalités d'une génétique d'association efficace et d'une sélection assistée par marqueurs dans le plan de croisement issu du programme de création variétale.

Dans cette thèse, nos efforts porteront sur deux caractères prioritaires qui vont permettre aux taillis de peuplier de remplir leur contrat de biomasse compétitive en quantité et en qualité. Au début de cette thèse, les connaissances biologiques et les moyens d'investigation disponibles sont différents pour ces deux caractères :

- **la résistance aux rouilles foliaires à *Melampsora larici-populina***. Cette maladie foliaire représente aujourd'hui la première contrainte biotique affectant significativement et quasi-annuellement les plantations de peuplier en Europe. Son développement est favorisé dans des milieux permettant le maintien d'une humidité élevée au niveau de la surface inférieure du feuillage. Les TCR et TtCR font partie de ces milieux. Le peuplier noir représente un important réservoir de variabilité génétique pour des sources de résistance vis à vis de cet agent pathogène avec lequel il co-évolue sur la majorité de son aire naturelle. Toutefois l'architecture génétique des résistances partielles présentes chez cette espèce est peu connue. Les efforts de recherche de régions génomiques et de gènes candidats associées à ces résistances ont jusqu'ici concerné les deux espèces *P. deltoides* et *P. trichocarpa* pour lesquelles les sources de résistances sont exaptées et se sont révélées fragiles (Villar et al. 1996; Cervera et al. 1996; Dowkiw and Bastien 2004; Jorge et al. 2005; Dowkiw and Bastien 2007; Dowkiw et al. 2010). Ces travaux ont permis toutefois de développer des méthodes de phénotypage fiables et à moyen débit permettant une

Aire de répartition naturelle

Population d'association
1100 individus

Panel de découverte de SNP
21 individus

Plan factoriel de croisement 5x4 *P. nigra*
937 individus

	SRZ	BDG	71077-308	92510-1
VGN	33	40	40	40
CBZ25	40	39	32	39
71041-3-402		38	12	40
71072-501	40	315	40	
SSC	39	39	33	38

Pédigrée de cartographie
329 individus F1

Figure 0.1 : Représentation schématique des ressources génétiques de peuplier noir disponibles et utilisées dans la thèse.

dissection de la résistance partielle observée en conditions naturelles d'infection à partir de tests réalisés en conditions d'inoculation contrôlée avec différentes souches de l'agent pathogène (Dowkiw et al. 2003). Dans cette thèse, je me suis intéressé à vérifier l'hypothèse d'un contrôle génétique et moléculaire complexe de la résistance partielle chez le peuplier noir avec pour objectif finalisé d'apprécier l'intérêt du développement d'une sélection assistée par marqueurs pour ce caractère.

- **les propriétés chimiques de la ressource ligno-cellulosique récoltée et le potentiel de saccharification**. Sachant que la production de bioéthanol est basée sur la conversion de la cellulose et des hémicelluloses et que la lignine représente un handicap pour le procédé industriel, il est important de préciser le niveau de variabilité génétique disponible et la nature du contrôle génétique de ces composants chez le peuplier noir. Ces caractères d'intérêt récent pour la sélection sont plus coûteux et difficiles à évaluer sur de larges collections de génotypes. Les premiers développements de modèles de prédiction indirecte de teneurs en composants chimiques du bois basés sur la spectrométrie proche infra-rouge sont prometteurs (Brinkmann et al. 2002; Robinson and Mansfield 2009). C'est pourquoi, ce travail de thèse s'est attaché à développer des modèles de calibration des composants chimiques du bois et du potentiel de saccharification chez le peuplier noir permettant une première détection des régions génomiques impliquées dans le contrôle génétique de ces caractères d'intérêt et l'identification de co-localisation éventuelle avec des gènes candidats déjà connus.

Les travaux réalisés dans cette thèse sont organisés dans les six chapitres suivants :

Chapitre I - *Comment la création variétale chez le Peuplier peut-elle répondre à des besoins accrus en ressource ligno-cellulosique dédiée à la production de biocarburants ?* Ce chapitre introductif a été conçu pour donner au lecteur le contexte agro-économique du sujet de thèse ainsi que les connaissances biologiques et génétiques sur l'espèce étudiée à savoir le peuplier noir. Il termine par une présentation des stratégies de sélection utilisées pour la populiculture classique et s'intéresse aux perspectives nouvelles offertes par les outils moléculaires pour maximiser l'obtention de gains génétiques à court terme pour la production d'une biomasse ligno-cellulosique de qualité.

Chapitre II – *Développement d'une boite à outils moléculaires pour la cartographie génétique et les études d'association chez P. nigra*. L'objectif de ce chapitre est l'identification de marqueurs SNP (Single Nucleotide polymorphism) dans une liste de gènes candidats choisis pour leurs liens avec les caractères cibles de sélection et leur couverture du génome de peuplier.

L'accès à des données de reséquençage de gènes entiers ou de fragments de gènes permettra d'analyser le patron de diversité moléculaire au sein du panel de découverte de 21 *P. nigra* et de préciser l'étendue du déséquilibre de liaison pour une meilleure estimation d'un taux de couverture en marqueurs optimal.

Chapitre III – *Développement de nouvelles cartes génétiques chez P. nigra*. Ce chapitre rapporte l'élaboration de cartes génétiques chez *P. nigra* à partir d'une famille F1 en ségrégation à l'aide d'une stratégie visant un ancrage de ces cartes sur la carte physique de *P. trichocarpa* et sur les différentes cartes génétiques déjà développées chez le peuplier afin de pouvoir comparer les régions génomiques détectées dans différents fonds génétiques.

Chapitre IV - *Cartographie de QTL associés à la résistance partielle à la rouille foliaire.* Ce chapitre comme le suivant s'intéresse à l'identification des régions génomiques impliquées dans la variabilité phénotypique de caractères d'intérêt en TCR et TtCR à l'aide d'une cartographie de QTL dans un pédigrée biparental mis en culture dans des conditions de TtCR et dont les cartes génétiques ont été élaborées dans le chapitre précédent. Les résultats relatifs aux deux caractères sont présentés successivement. Dans ce chapitre, la détection de QTL associés à la résistance à la rouille foliaire est présentée sous forme d'un manuscrit en anglais soumis à publication à la revue Plos One.

Chapitre V - *Cartographie de QTL associés à la qualité de la biomasse*. Outre la détection de QTL associés plus spécifiquement aux composants chimiques du bois et au potentiel de saccharification, sont présentés dans ce chapitre les modèles de calibration SPIR développés pour ces différents caractères.

Les conclusions et perspectives de l'ensemble de ces travaux font l'objet du **chapitre VI** de cette thèse.

Chapitre I

Introduction

- Comment la création variétale chez le Peuplier peut-elle répondre à des besoins accrus en ressource ligno-cellulosique dédiée à la production de biocarburants ?

Chapitre I : Comment la création variétale chez le Peuplier peut-elle répondre à des besoins accrus en ressource ligno-cellulosique dédiée à la production de biocarburants ?

Dans le souci de répondre à une demande énergétique qui ne cesse d'augmenter et une raréfaction des ressources actuelles en énergie, plusieurs projets de recherche ont vu le jour pour développer des alternatives aux sources d'énergie classique. Les ressources pétrolières n'étant pas inépuisables, il est en effet important de gérer au mieux les réserves précieuses de combustibles fossiles pour les générations futures (un principe fondamental du développement durable) mais également de réduire les émissions mondiales de gaz à effet de serre, en particulier le dioxyde de carbone (CO_2). À ce titre, la biomasse végétale constitue une source d'énergie neutre car la quantité de CO_2 émise lors de la combustion correspond à celle qui a été nécessaire à la croissance des plantes à l'origine de la biomasse produite. La production de combustibles ou de produits chimiques à partir de la biomasse végétale n'est pas un nouveau concept. En effet la cellulose, les huiles végétales, l'éthanol ou le méthanol et d'autres composés préparés à partir de végétaux ont été utilisés pendant plus de deux siècles pour préparer des solvants, des peintures, des adhésifs ou encore des textiles synthétiques. À partir des années 1960, cette filière s'est vue supplantée par l'industrie pétrochimique. La crise pétrolière des années 1970 a ensuite suscité un nouvel intérêt pour ces bioressources, intérêt qui s'est estompé rapidement avec la chute du prix du pétrole. De nos jours, le contexte économique, géopolitique et environnemental est de nouveau favorable à l'exploitation de la biomasse végétale à des fins énergétiques. Selon l'association Prospective 2100 qui réunit plusieurs industriels et organismes de recherche (Sofiprotéol, Dalkia, Procéthol, CEA …) dans le domaine de biomasse et énergies, la consommation mondiale énergétique devrait doubler d'ici à 2050, entrainant une augmentation de la production énergétique mondiale à base de biomasse végétale qui passerait de 1,1 à 2,3 x 10^9 tonne équivalent pétrole par an (tep/an)

A. Valorisation de la ressource ligno-cellulosique pour la production de biocarburants

1. Un contexte en faveur de la chimie verte et d'une utilisation accrue du bois comme source d'énergie

En écologie, on définit la **biomasse** comme la masse totale des organismes vivants (animaux/végétaux) mesurée dans un environnement. La biomasse peut être aussi définie comme l'ensemble des matières organiques pouvant devenir des sources d'énergie. On distingue trois sortes de **biomasse végétale**, une biomasse ligno-cellulosique (bois, paille, résidus verts, etc.),

une biomasse riche en sucres et en amidon (betterave, canne à sucre, blé, maïs, etc.) et une biomasse oléagineuse (colza, soja, tournesol, etc.). La chimie verte est notamment la chimie des nouvelles énergies, des nouveaux produits et matériaux issus de la biomasse végétale. Mais pour se substituer à la pétrochimie, cette chimie verte doit encore démontrer son efficacité industrielle, économique et environnementale. Dans ce domaine, innovations et évaluations progressent. Actuellement, la biomasse végétale est valorisée sous forme de combustibles, biogaz et biocarburants (d'après l'observatoire des énergies renouvelables (http://www.energies-renouvelables.org) et constitue une matière première de produits chimiques (Ragauskas et al. 2006).

Les combustibles végétaux dont le bois énergie :

Sont utilisés en combustion, le bois massif, les déchets de bois mais également d'autres matières organiques telles que la paille, les résidus solides des récoltes, les rafles de maïs, la bagasse de la canne à sucre, les grignons d'olives, etc. En France, comme dans la plupart des pays européens, le prélèvement forestier reste inférieur à l'accroissement naturel de la forêt, le bilan carbone est donc positif. De 2005 à 2010, le ratio Prélèvement/(Production – Mortalité) était inférieur à 50% (IFN 2011). Il existe aujourd'hui des appareils à combustible bois innovants et efficaces à disposition des particuliers comme des collectivités ou des industries. Les chaudières à biomasse brûlent différents biocombustibles : granulés de bois, bûches, plaquettes forestières, sciures ou copeaux.

En 2010, 64% de la production totale d'énergie renouvelable thermique était d'origine bois énergie, ce qui représentait 99% de l'utilisation du bois énergie, tandis que la production d'électricité du bois énergie représentait 2% de la production totale d'énergie renouvelable électrique et 1% d'utilisation du bois énergie (FCBA 2012). Le cabinet Xerfi (expert en économie) a pointé le manque d'investissement actuel dans cette filière, et met en cause la difficulté de fiabiliser l'approvisionnement en bois et donc d'établir des contrats à long terme, rédhibitoire auprès des investisseurs. Les centrales électriques au bois exigent notamment davantage de capitaux que l'éolien ou le solaire, pour un retour sur investissement moins rapide. Xerfi considère qu'au-delà de ces obstacles, la biomasse végétale reste pourtant l'une des sources énergétiques les plus attractives, car la ressource est moins aléatoire que le vent par exemple.

L'Ademe (Agence de l'Environnement et de la Maîtrise de l'Énergie), dans le cadre de son second appel à projets intitulé « Biomasse Chaleur Industrie, Agriculture et tertiaire » destinés aux acteurs de la vie économique, vise pour 2012 la production de 175 000 tep pour limiter la

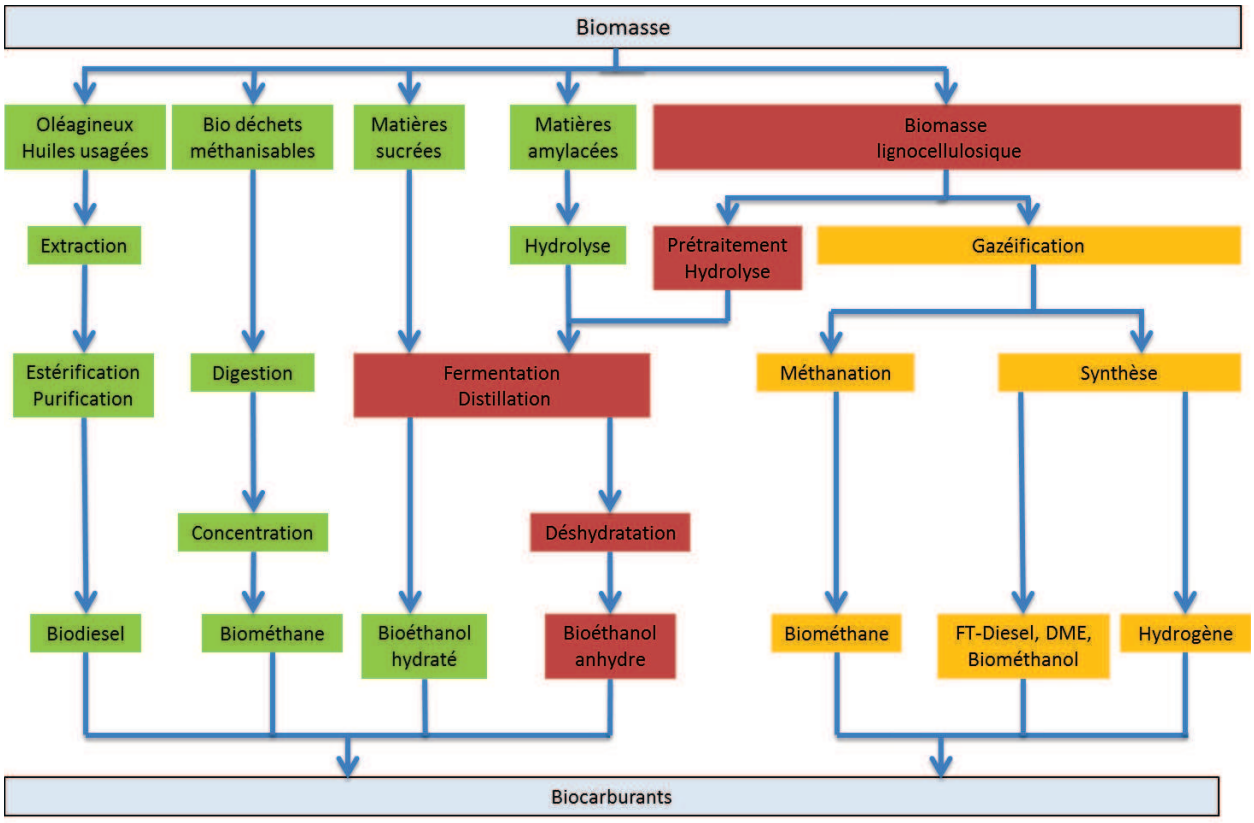


Figure I.1 : Les filières de valorisation énergétique de la biomasse en biocarburants. En rouge bordeaux, les ressources ligno-cellulosiques et leur parcours de transformation pour la production de bioéthanol dans le cadre du projet Futurol. D'après http://www.plateforme-biocarburants.ch

dépendance à la hausse du coût des énergies fossiles, garantir l'approvisionnement énergétique à long terme et réduire les impacts environnementaux et les coûts de fonctionnement. La tep vaut, par définition, 41,868 x 109 Joules, ce qui représente l'énergie dégagée par la combustion d'une tonne de pétrole. Elle sert notamment à comparer entre elles les différentes sources d'énergie. 1 tep = 11,628 MWh d'électricité = 1 616 kg de houille = 1 069 m3 de gaz d'Algérie = 954 kg d'essence moteur d'après le site du ministère de l'écologie, du développement durable et de l'énergie ; une tonne de bois anhydre (tMS :tonne Matière Sèche) équivaut à 5,06 MWh soit 0,43 tep, (FCBA 2009b).

Biocarburants

Les biocarburants de première génération sont principalement commercialisés sous trois formes au niveau mondial : le bioéthanol, le biodiesel et le biogaz, même si cette dernière reste encore à l'heure actuelle moins développée que les deux précédentes dans le domaine des transports (**Figure I.1**).

Bioéthanol

Le bioéthanol est en fait de l'alcool éthylique (ou encore éthanol), identique par sa composition à l'alcool de bouche. Tous les sucres fermentescibles (glucose, saccharose, etc.) peuvent être transformés en éthanol par fermentation. Ces sucres sont présents dans un état plus ou moins polymérisé dans de nombreuses espèces du monde végétal comme la betterave à sucre, la canne à sucre, le blé, le maïs, la pomme de terre, mais également dans les fourrages ou encore le bois. Des déchets tels que le petit lait ou le vieux papier peuvent également être transformés en bioéthanol. Suivant l'état de polymérisation, ces sucres doivent subir un (ou plusieurs) traitement(s) préalable(s), dont le but est de transformer les chaînes de polymères en sucres simples. Après fermentation à l'aide de micro-organismes (levures, bactéries, etc.), l'éthanol est recouvré par distillation (éthanol hydraté à 95-96% v/v), puis par déshydratation (éthanol anhydre à 99,7% v/v).

Biodiesel

Le biodiesel est un ester (généralement ester méthylique) habituellement issu d'huile végétale. L'utilisation d'huile végétale comme carburant n'est pas une technique nouvelle. Rudolf Diesel, à l'origine, avait en effet développé le moteur qui porte son nom à l'aide d'huile d'arachide. Les huiles végétales s'obtiennent classiquement par simple pressage de graines oléagineuses telles que le colza, le tournesol, le soja, etc. Des graisses animales ainsi que des huiles alimentaires usagées peuvent également être utilisées à cet effet. Toutes ces huiles offrent

cependant une trop forte viscosité et un indice de cétane (aptitude à l'auto-inflammation) trop faible, rendant difficile leur utilisation directe dans un moteur diesel traditionnel. L'IFPEN (Institut Français du Pétrole Énergies Nouvelles) fut pionnier dans le développement des biocarburants de première génération en France. Il est à l'origine du procédé Esterfip-H ^{TM} pour la production de biodiesel.

Biogaz

La digestion anaérobie ou méthanisation consiste à transformer les matières organiques (sucres, protéines et graisses) en l'absence d'oxygène pour produire du biogaz, à savoir un mélange de méthane (CH_4) et de dioxyde de carbone (CO_2), au moyen de bactéries. Le biogaz est généralement utilisé comme combustible pour la production locale de chaleur ou la cogénération d'électricité et de chaleur. Toutefois, l'utilisation du biogaz comme carburant dans des véhicules à gaz se développe de plus en plus (ADEME et al. 2008).

Les biocarburants de première génération ont fait l'objet de plusieurs débats. Ils sont pointés du doigt du fait des conflits d'usage des terres qu'ils engendrent à l'échelle mondiale. Ils sont accusés d'être à l'origine d'un déboisement important en zones tropicales pour mettre à disposition de l'agriculture de nouvelles surfaces. En ce sens, ils réduisent la capacité de stockage à long terme du CO_2 et menacent la préservation de la biodiversité (Persillet 2012; Searchinger et al. 2008; ADEME 2010). Ils ont également engendré une compétition entre cultures à usage alimentaire et cultures à usage non-alimentaire se traduisant notamment par une flambée des cours de denrées agricoles.

Les politiques gouvernementales reportent aujourd'hui leur intérêt sur les biocarburants dits de deuxième génération qui utilisent cette fois la plante entière, les feuilles, les pailles ou les tiges des productions alimentaires ou des plantes spécifiquement dédiées à la production de biocarburants comme le miscanthus. Priorité est donnée à ce que ces biocarburants ne mettent pas en concurrence sécurité énergétique et sécurité alimentaire.

Ces biocarburants de deuxième génération font l'objet d'une activité intensive de recherche et développement (R&D) depuis plusieurs années et représentent pour beaucoup la filière la plus prometteuse à moyen-long terme. Il n'existe toutefois à ce jour aucune installation industrielle de ce type. Deux voies de production à partir de la ressource ligno-cellulosique sont principalement étudiées : la conversion thermochimique et la conversion biochimique.

La **conversion thermochimique** implique essentiellement la gazéification de la biomasse, ou autre procédé thermique, suivie de la synthèse catalytique ou de la fermentation du gaz ou du liquide qui en résulte. **La conversion biochimique** implique essentiellement l'hydrolyse des

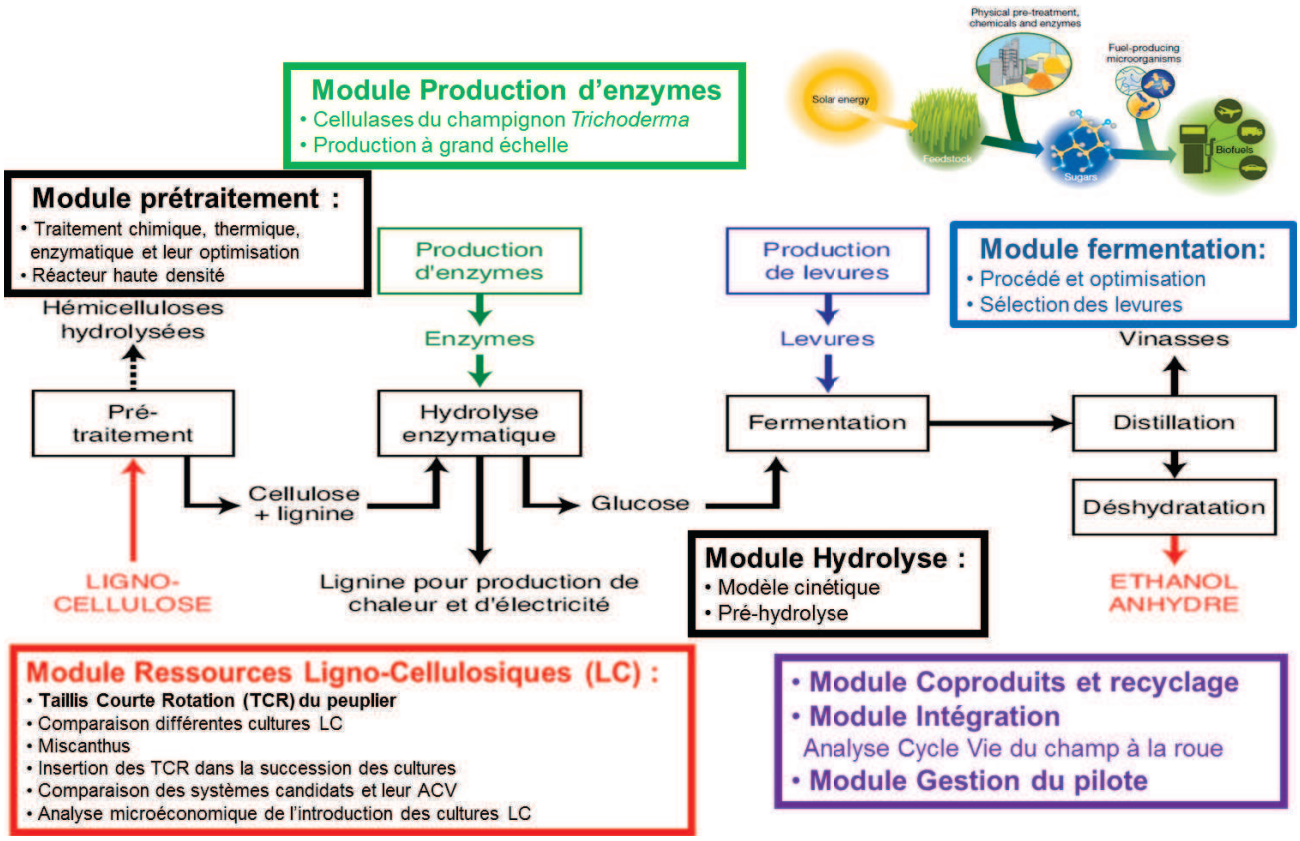


Figure I.2 : Les différents modules du projet Futurol avec les actions à accomplir. Le travail de thèse intervient en amont du processus industriel dans le cadre des cultures dédiées et plus spécialement, les taillis à courte rotation et à très courte rotation du peuplier (http://projet-futurol.com)

polysaccharides de la biomasse, et la fermentation des sucres qui en résultent en éthanol. Les carburants produits à partir de biomasse ligno-cellulosique sont communément appelés carburants cellulosiques.

C'est dans le cadre de la **conversion biochimique** que le projet Futurol a vu le jour en 2008 (**Figure I.2**). Il est porté par la société Procéthol 2G avec 11 acteurs majeurs (ARD, IFPEN, INRA, Lesaffre, Champagne Céréales, ONF, Tereos, Total, Crédit Agricole Nord Est, CGB, Unigrains). Il inclut la création d'une unité pilote de production sur le site de Pomacle-Bazancourt dans la Marne. Inaugurée en octobre 2011, cette unité de production a une capacité de production de 180000 litres de bioéthanol par an. La biomasse végétale utilisée est diversifiée : plantes dédiées (TCR ou TtCR de peuplier, panic érigé ou Switchgrass et miscanthus), bois, coproduits agricoles (paille de céréales, pulpes de betteraves…), résidus forestiers et déchets verts urbains. En 2015, le projet prévoit la mise en marche d'un prototype préindustriel d'une capacité de 3,5 millions de litres de bioéthanol avant de mettre en production à partir de 2016 une unité industrielle ayant une capacité de production de 180 millions de litres de bioéthanol par an.

2. Quelle production de bois pour la production de biocarburants ?

En France, la biomasse végétale représente actuellement 4,6% de l'énergie primaire consommée, ce qui correspond aux deux tiers des énergies renouvelables (Association « Prospective 2010 »). Il semble possible de tripler cette production d'énergie à partir de biomasse végétale grâce à une intensification de la production forestière compatible avec une gestion durable des écosystèmes, à une meilleure valorisation des sous-produits de cultures et en réduisant les exportations pour valoriser les produits sur place. Afin d'atteindre une production de 35 à 40 Mtep/an (contre 12.5Mtep/an aujourd'hui), il faudrait également mobiliser 5 millions d'hectares cultivés sous forme de cultures dédiées dont les deux tiers seraient consacrés aux biocarburants dont la production passerait de 2 Mtep/an en 2010 à 10 Mtep/an en 2050. La France devrait donc être en mesure d'atteindre 20% d'énergies renouvelables dans son bouquet énergétique comme prévu par l'Union européenne (EU-Commission 2008).

En 2010, la consommation mondiale de biocarburants de première génération a représenté 3 % de la consommation totale de carburants, soit 55 Mtep. Parmi ces biocarburants on retrouve 73 % de bioéthanol utilisable dans les motorisations essence, et 27 % de biodiesel pour les motorisations diesel (http://www.ifpenergiesnouvelles.fr, site internet de l'IFPEN).

Pour atteindre ces enjeux ambitieux pour la production de bioéthanol ligno-cellulosique 2ème génération (Lorne and Chabrelie 2011), il convient de travailler simultanément sur différents fronts:

- Augmenter la mise à disposition d'une ressource ligno-cellulosique de qualité en particulier à travers le développement de cultures dédiées. Ces cultures dédiées (TCR ou TtCR, cultures de miscanthus ou switchgrass) sont des cultures nouvelles, peu connues des exploitants agricoles. D'importants travaux de recherche et développement sont nécessaires pour, 1) permettre d'adapter ces cultures aux contextes locaux (climat, sols, économie des exploitations, etc.), 2) préciser les itinéraires techniques (désherbage, récolte, optimisation, logistique, etc.), 3) définir les caractéristiques du matériel végétal souhaité (idéotype) et envisager de nouveaux programmes de sélection, et 4) assurer l'acceptabilité sociale de ces nouvelles cultures (impacts sur les paysages notamment).
- Mettre au point des procédés de broyage, de prétraitement et de saccharification économiquement viables pour libérer les sucres simples contenus dans cette ressource ligno-cellulosique. Ces sucres simples sont des sucres à squelette de 6 carbones (C6) ou encore des sucres à 5 carbones (C5) (Rubin 2008). Parmi les différents procédés de prétraitement visant à libérer les sucres de la plante (Zhu et al. 2010), il reste à déterminer le procédé le plus efficace à une échelle industrielle, et qui soit suffisamment flexible pour une ressource biomasse variée. Par ailleurs, le coût de production des enzymes pour l'étape d'hydrolyse des sucres est également un facteur limitant. De récentes annonces d'industriels producteurs d'enzymes présagent des progrès importants avec des réductions significatives de coûts.
- Toujours au niveau industriel, identifier les micro-organismes possédant le patrimoine génétique permettant de fermenter un maximum de ces sucres. Les levures ne sont pas capables de fermenter la totalité des sucres libérés par la biomasse cellulosique et surtout les sucres C5. Des travaux en unités pilotes ont montré l'efficacité de certaines souches sur les sucres C5. La productivité de cette fermentation reste cependant améliorable.

Il est également important d'appréhender dès aujourd'hui quels pourront être les impacts socioéconomiques et environnementaux du développement de ces nouvelles filières.

B. Potentiel des taillis à courte rotation d'arbres

1. État de l'art, Agronomie et cultures dédiées (toutes espèces dont peuplier)

La ressource ligno-cellulosique issue des taillis à courte rotation (TCR) ou à très courte rotation (TtCR) représente l'une des sources d'énergie qui pourrait avoir un intérêt pour la

réduction de gaz à effet de serre et la sécurité de l'approvisionnement en énergie (Perttu 1998; Britt et al. 2002 ; Caslin et al. 2010). Parfois, les anglophones appellent les TCR, les SRF (Short Rotation Forestry), tandis que TCR et TtCR sont regroupés sous l'acronyme SRC (Short Rotation Coppice). Le concept de TtCR s'est beaucoup développé depuis les années 1980 dans les pays scandinaves et anglo-saxons, avec le double objectif de réaliser un traitement tertiaire des eaux usées et de produire de l'énergie avec la biomasse produite.

Actuellement, les TCR et TtCR sont cultivés exclusivement pour des usages industriels (papeteries) ou énergétiques (chaufferies bois), leur utilisation pour la production des biocarburants 2ème génération est en cours de développement. Ces peuplements nécessitent des terrains à forte potentialité agronomique, profonds, humifères et bien irrigués et des espèces ligneuses à fort potentiel de croissance et rejetant de souche. Outre leur rôle énergétique, les TCR et TtCR peuvent s'inscrire dans une perspective d'aménagement du territoire, de dépollution des sols et des boues d'épuration (phytoépuration et phytoremédiation) (Dimitriou and Rosenqvist 2011), et de protection des espaces agricoles abandonnés ou gelés (jachères).

À mi-chemin entre la foresterie traditionnelle et l'agriculture, les TCR et TtCR sont des cultures intensives d'arbres plantés à haute densité, exploités selon un cycle court de 2 à 10 ans et qui rejettent de souche assurant une production soutenue sur plusieurs rotations. Au cours de l'hiver suivant la plantation, la tige principale est coupée à 10 centimètres du sol pour stimuler la production de plusieurs rejets par souche : c'est le recépage. Plusieurs critères de l'itinéraire de culture peuvent influencer la productivité des taillis. Le choix du matériel végétal est important et doit se faire notamment en fonction du site de plantation. L'hétérogénéité des sols affecte le rendement des taillis ; les forts rendements sont obtenus sur sols organiques ; le traitement des adventices est aussi important pour ce genre de culture et demande un certain développement technique ; la fertilisation au-dessous des recommandations peut réduire le rendement de 20% et l'effet des fertilisants est dépendant du désherbage. La disponibilité d'eau est un facteur limitant, les demandes en eau sont très grandes. La récolte se fait généralement en hiver quand les feuilles sont tombées, le taux d'humidité dans le bois est le moins élevé et la quantité des nutriments prélevés est minime afin de maintenir la viabilité de la plantation et de favoriser le rejet de souche (Kauter et al. 2003). Deux techniques de récolte sont possibles : la technique de passage unique pour la coupe et le broyage (Single pass 'cut and chip') et la récolte en tiges entières (Whole stem) (Mitchell et al. 1999). Dans le cas où on stocke les plaquettes fraichement coupées (teneur en eau correspondant à 50% à 60% de la biomasse fraiche) sans ventilation, ni séchage, le développement de champignons entrainera la perte de biomasse de 3% par mois (Mitchell 1995).

Principales espèces forestières utilisées en TCR et en TtCR :

En Europe, peu d'espèces forestières sont appropriées pour les cultures énergétiques. Elles sont traditionnellement plantées, comme le saule, le peuplier, l'eucalyptus, l'aulne, le frêne, le bouleau, et le robinier faux acacia, à des fins différentes (industrie de la pâte, paniers, etc.) (Petersen 2007). Dans la littérature, 5 espèces ou groupes d'espèces sont majoritaires : les saules et les peupliers qui appartiennent à la famille des Salicacées, les eucalyptus, l'aulne et le robinier faux acacia (**Tableau I.1**).

Les Salicacées

Le genre *Salix* :

Les saules sont plantés au nord de l'Europe en TtCR. Avec 16 000 hectares de taillis de saules, la Suède est le pays où les TtCR de saules sont les plus développés. La densité de plantation des TtCR est très forte (10 000 à 15 000 tiges /ha en moyenne). On récolte tous les 2 à 4 ans, en hiver, des petites tiges directement transformées en broyat valorisé en biocombustible. Il ressort que les rendements entre parcelles varient entre 6 et 12 tonnes de MS/ha/an. Les saules cultivés en TtCR pour la production de biomasse ont fait l'objet d'une sélection variétale de clones performants pour ce type d'usage (Karp et al. 2011).

Le genre *Populus* :

À l'opposé de la populiculture classique où la densité de plantation est de 150 à 200 plants/ha, la culture du peuplier en taillis se caractérise par une densité de plantation élevée oscillant entre 1000 et 3000 tiges/hectare pour les TCR et entre 8000 à 15000 tiges/ha pour les TtCR (Petersen 2007 ; FCBA 2009a). Les deux systèmes de culture visent la production de produits différents. Une densité autour de 2000 tiges/hectare en TCR semble être un bon compromis pour obtenir à 7 ans des tiges de 12 à 16 cm de diamètre permettant la préparation de billons de 1m et 2m destinés jusqu'ici majoritairement à l'industrie de la trituration ou au chauffage domestique (FCBA 2009a). Les très fortes densités utilisées en TtCR visent à une production de biomasse élevée sur plusieurs tiges broyées lors d'une récolte mécanisée. Les clones plantés sont issus de programmes de sélection européens destinés à la populiculture traditionnelle (Berthelot and Gavaland 2007).

Les espèces fixatrices d'azote : Aulnes et *Robinia pseudo acacia*

Les aulnes peuvent être multipliés par reproduction végétative à partir des tiges semi-ligneuses et rejettent de souche. Grâce aux mycorhizes contenus dans les nodosités que portent leurs racines les aulnes enrichissent les sols pauvres en nitrates en fixant l'azote de l'air (une

Tableau I.1 Productivité moyenne en taillis à courte rotation (TCR) et très courte rotation (TtCR) de 5 espèces forestières majoritairement utilisées en zones tempérées. ES : Espagne ; PT : Portugal ; FR : France ; UE : Union Européenne ; Ro : Roumanie ; Su : Suède ; RU : Royaume Uni ; PL : Pologne ; HU : Hongrie ; IT : Italie.

Espèce	Usage TCR ou TtCR	Rotation moyenne (ans)	Productivité moyenne t MS/ ha/an	Zones de production		Programmes de sélection	Source
				Pays	Surfaces (ha)		
Eucalyptus	TCR	9 - 12	9 - 18	ES, PT, FR	1 200 500	oui; clones	(Melun 2008) (Aronsson et al. 2006)
	TtCR	3 - 7	10				
Aulne	TCR	20	5			non; graines	(Petersen 2007)
	TtCR	2 - 4	0,15 - 8				(Gerstch et al. 2012)
Peupliers	TCR	6 - 14	4 - 9	UE	700	oui; clones	(Petersen 2007) (Gerstch et al. 2012)
	TtCR	3 - 4	2 - 18	UE	370 000		(Ball et al. 2005) (Gerstch et al. 2012)
Saule	TtCR	2 - 4	6 - 12	Ro, Su, RU, PL	42 175	oui; clones	(Weih et al. 2006; Weih and Bonosi 2009)
Robinier faux acacia	TtCR	2-4	3 - 10	HU, IT	>500	oui; graines	(Grunewald et al. 2009)

symbiose avec une bactérie actinomycète du genre *Frankia*). Ces caractéristiques vont permettre à cette espèce d'être utilisée en TtCR (Jorgensen et al. 2005) pour leur action fertilisante à des fins d'agriculture biologique, de régénération des terrains agricoles abandonnés (Uri et al. 2009; Uri et al. 2011) ou de production de biomasse en TCR avec une rotation plus longue en valorisant des essences forestières locales (Bogdan et al. 2009).

Le *Robinia pseudo acacia* est une essence de lumière qui résiste bien au froid par suite de sa mise en végétation tardive, il s'accommode de sols pauvres grâce aux bactériorhizes qui lui permettent de fixer l'azote atmosphérique dans le sol. Grâce à cette capacité, le robinier faux acacia est plus productif en biomasse que le peuplier et le saule sur des terrains post-miniers (Grunewald et al. 2009). La multiplication de cette espèce se fait par graines, drageons et greffes pour les cultivars. Son bilan énergétique a été évalué en TCR. Il est positif puisque sa production de biomasse convertible en bioéthanol ne demande que peu d'apports en fertilisants et en irrigation (Redei et al. 2009 ; Gonzalez-Garcia et al. 2011). Il existe des programmes de sélection de cette essence en Hongrie (Redei 1998; DeGomez and Wagner 2001) et sa propagation se fait majoritairement par graines.

Les Eucalyptus :

Environ 1 million d'hectares sont plantés au Portugal et en Espagne (Melun 2008). Les différentes espèces restent très sensibles au froid ce qui réduit leur implantation en France à des régions du Midi et du Sud-Ouest : 2000 hectares y ont été installés à partir des années 1980 (Melun 2008); Néanmoins, une étude montre l'intérêt de cette essence pour la culture en taillis. Les résultats montrent des taux de croissance élevés et des rendements énergétiques calorimétriques élevés (Dalianis et al. 1994). Des programmes d'amélioration génétique sont actuellement développés en France au FCBA et de nouvelles sorties variétales viendront compléter l'offre de matériel végétal dans les années à venir (Melun 2008).

Productivité des TCR et TtCR :

À part la Suède, peu de pays possèdent des plantations commerciales de grande taille de TCR ou TtCR dédiées pour l'énergie. Le **tableau I.1** résume les productivités de TCR et de TtCR reportées en Europe. Cette productivité varie selon l'espèce, la zone de plantation et les itinéraires techniques utilisés. Une synthèse s'intéressant aux TtCR à des fins de phytoremédiation montre à partir de 29 publications différentes une variabilité du rendement des TtCR des cinq espèces utilisées pour des fins énergétiques (Gerstch et al. 2012).

Cette variabilité de niveau de production s'explique par plusieurs facteurs. Le niveau de production dépend du **choix du site de plantation**, un sol pauvre et acide donnant des rendements très faibles même avec du matériel végétal à fort potentiel productif (Kauter et al. 2003). Il est recommandé d'installer les TCR et TtCR dans des sols profonds d'au moins 1m (Zalesny and Bauer 2007).

Différentes composantes de l'itinéraire technique influencent significativement les productivités observées :

- La préparation du site. Elle est essentielle pour la réussite de l'établissement d'une plantation de saule (Abrahamson et al. 2002) ou du peuplier (Mitchell et al. 1999 ; Stanturf et al. 2001). La végétation existante doit être éliminée à l'aide d'herbicide et le sol doit être labouré profondément, hersé et laissé nu durant l'hiver pour permettre au gel de fractionner la structure du sol.
- La densité de plantation. Une comparaison de différentes densités de plantation entre des taillis de *P. xcanadensis*, *P. deltoides* et saules montre que le rendement en biomasse augmente avec la densité de plantation pour atteindre son maximum à 10000 plants/ha en Italie (Bergante et al. 2010). Cependant, des taux de mortalité élevés sont signalés à des fortes densités de plantations : environ 15 % à une densité de 110000 plants/ha de saule (Britt et al. 2002) et 20% de mortalité pour des plantations de *P. deltoides* à 10000 plants/ha (Bergante et al. 2010).
- L'entretien après plantation. Malgré le fait que les espèces utilisées en taillis sont des espèces pionnières caractérisées par une croissance juvénile rapide qui permet une fermeture rapide du couvert et donc limite la disponibilité de lumière pour les plantes adventices compétitrices (FAO 1979), le désherbage reste un élément clé pour la réussite de l'installation des taillis (Buhler et al. 1998; Mitchell et al. 1999; Dillen et al. 2011 ; Broeckx et al. 2012).
- La fertilisation. La fertilisation pour compenser les exportations dues aux retraits réguliers de la biomasse (Dimitriou and Rosenqvist 2011) constitue aussi un élément important pour avoir de bons rendements.
- Le choix du matériel végétal (choix d'espèces et choix de la variété) et son adéquation avec le site de plantation et l'itinéraire sylvicole. Ainsi, l'utilisation du robinier faux acacia sur des sites ex-miniers a donné un meilleur rendement en comparaison avec *Salix* et *Populus* réputés plus productifs (Grunewald et al. 2009). À différentes densités de plantations (10000 à 111000 plants/ha) *S. viminalis* qui produit des tiges dressées supporte mieux des densités élevées que *S. dasyclados* qui a des tiges plus étalées (Bullard et al. 2002). En plantation monoclonale, les travaux sur peuplier (Heilman and Stettler 1986; Al-Afas et al. 2007; FCBA 2009a ; Bergante et al. 2010; Benomar et al. 2012; Broeckx et al. 2012), et sur saule (Stolarski et al. 2008;

Wilkinson et al. 2007) ou de comparaison entre peuplier et saule (Labrecque and Teodorescu 2005; Aylott et al. 2008) et sur le robinier faux acacia (Redei et al. 2009), montrent tous qu'il y a un effet important du clone sur le rendement ainsi que de fortes interactions clone x densité de plantation. Ainsi, le clone MB915 (*P. maximowiczii* x *P. balsamifera*) donne des rendements plus élevés à forte densité (1m x 1m d'espacement) qu'à faible densité (5m x 5m) (Benomar et al. 2012).

- La durée de rotation. La productivité maximale de la plantation est atteinte quand l'accroissement annuel moyen (MAI = rendement total à la récolte/durée de la rotation) atteint son maximum (Bonduelle 1989). Cet accroissement permet de déterminer la durée de la rotation optimale pour maximiser les rendements. Généralement, on récolte les arbres avant que le MAI maximal ne soit pas atteint pour des raisons de coût puisque le coût de récolte qui représente 70 % du coût total des taillis (Mitchell 1995) augmente considérablement avec le diamètre des arbres récoltés. Des récoltes moins fréquentes permettraient donc de réduire le circuit unitaire de la production de bois et d'optimiser le rendement (Strong 1989). De plus, des récoltes répétées peuvent entrainer une dégradation du sol et un stress abiotique des arbres (Mitchell et al. 1999) qui réduira la productivité des arbres par la suite.

2. Exigences et limites actuelles des systèmes de culture en taillis pour la bioénergie

Différents obstacles techniques et non techniques au développement des cultures énergétiques ont été publiés par l'IEA (International Energy Agency) (Alker et al. 2005; Petersen 2007) (**Tableau I.2**). Les barrières techniques peuvent être regroupées en différentes catégories s'adressant respectivement aux itinéraires techniques, à la disponibilité en matériel végétal, aux processus industriels et à la valeur technologique des produits, l'organisation de la filière et son économie. Si les activités de recherche et développement vont permettre de lever les barrières techniques, les barrières non techniques reliées au volet social, économique, environnemental et politique seront plus complexes à résoudre et demandent une volonté politique nationale. Souvent les barrières techniques et non techniques se croisent, par exemple, les technologies de conversion de la biomasse doivent démontrer leur performance pour attirer les investissements (Alker et al. 2005).

Parmi les connaissances à acquérir, il est important de mieux caractériser la valeur de la biomasse produite, d'identifier les possibilités d'amélioration génétique en prenant soin de sélectionner un matériel adapté aux itinéraires sylvicoles et aux sites de production choisis

Tableau I.2 : Évaluation des obstacles techniques et non techniques des taillis à courte rotation dédiés pour la bioénergie (d'après Petersen, 2007)

Obstacles techniques		Obstacles non techniques
Itinéraires sylvicoles	Des connaissances réduites des interactions sol-site-productivité Un épuisement éventuel des éléments nutritifs du sol Un épuisement éventuel des ressources en eau Des inquiétudes sur la biodiversité et des effets sur la faune Les inquiétudes sur la nature invasive des cultures Une augmentation des pressions environnementales dues à la circulation Un matériel de récolte peu développé	Un manque d'innovation ou de communication d'idées novatrices Une politique gouvernementale pas claire et incohérente Un manque et une mauvaise utilisation des fonds publics pour la R & D Une politique ne reflétant pas les avantages multiples des taillis et une politique d'urbanisation trop restrictive Les biocarburants ne sont pas compétitifs par rapport aux carburants fossiles entrainant une faible confiance des investisseurs Peu de personnel qualifié dans le secteur de la production des biocarburants Une chaîne d'approvisionnement sous-développée Une concurrence avec les combustibles fossiles pour le développement de technologies de conversion Des craintes si les taillis deviennent économiquement intéressants, les cultures pour la nourriture deviennent menacées (concurrence avec la production d'autres cultures) et une crainte vis-à-vis des forêts indigènes qui risquent de disparaitre au profit des taillis Un manque de soutien pour le développement des systèmes de production et peu de financement pour des technologies de pointe Des partenariats entre diverses industries avec des expériences différentes et peu de personnel qualifié dans les technologies de conversion Un développement des monopoles Une mauvaise perception par les gens de l'énergie des taillis, un malentendu public sur la biomasse et un manque de sensibilisation et des idées fausses au sujet de la culture Des subventions sur les carburants fossiles ou ceux de la biomasse ont un effet négatif sur la compétitivité des taillis Un rejet ou une adoption lente de ces taillis comme matière première pour la production d'énergie Une discordance entre la nature transitoire de la politique et la nature à long terme de la récolte Des solutions non-renouvelables ne sont pas souvent le premier choix Une compétition des taillis avec les déchets et résidus de la biomasse une mauvaise communication entre les différents secteurs de la filiale
Disponibilité du matériel végétal	Un manque dans les programmes nationaux de sélection Une compréhension insuffisante de la combinaison optimale des variétés Une disponibilité limitée de plants	
Processus industriel et valeur technologique des produits	Peu de connaissances sur les propriétés des matériaux combustibles issus des taillis Une mauvaise performance des technologies de pointe Des incohérences dans la qualité du carburant provoquant des mauvaises performances pour les technologies conventionnelles Préoccupation des émissions atmosphériques	
Organisation de la filière	Une discordance dans la constance entre l'approvisionnement et la demande Des rendements non rentables	
Économie	Des coûts de transport élevés en raison des caractéristiques du carburant produit Un bilan énergétique négatif Une Période trop longue entre l'investissement et le retour de cet investissement	

(**Tableau I.2**). Les coûts d'installation des taillis doivent être faibles pour permettre une gestion à long-terme et un retour de l'investissement favorable.

En tout cas, pour la réussite des TCR et TtCR énergétiques, il faut déployer beaucoup de moyens pour la communication avec les agriculteurs en général et ceux qui souhaitent se convertir en agro-sylviculteurs et que les projets pilotes de production de biocarburants à partir des taillis, comme le projet Futurol, démontrent leur bilan favorable de conversion pour attirer les investissements (Luger 2000).

Le projet FUTUROL qui vise le développement et la commercialisation d'un procédé complet de production de bioéthanol de 2ème génération à partir de plantes entières ou de biomasse lignocellulosique vise à surmonter un certain nombre de ces obstacles. Il souhaite développer un procédé durable intégrant une grande variété de matières premières et qui soit adaptable notamment en fonction de la zone géographique où il sera mis en œuvre (territoire, climat) mais également selon les saisons. L'objectif du projet est de mettre sur le marché un procédé, des technologies et des produits (enzymes et levures) permettant d'atteindre les résultats suivants :

1/ Produire du bioéthanol à un **prix compétitif** grâce à une matière première diversifiée (coproduits agricoles, biomasse forestière, cultures dédiées, …)

2/ Développer des technologies d'extraction de la cellulose, sélectionner des enzymes et des levures et mettre au point des procédés d'hydrolyse et de fermentation les mieux **adaptés** à chaque configuration de matières premières.

3/ Obtenir les meilleurs **bilans énergétiques** et de **gaz à effet de serre** possibles sur l'ensemble de la chaîne de production.

4/ S'inscrire dans une logique de **développement durable** sur le long terme et tout au long de la filière du **champ à la roue**.

3. Nouveaux objectifs d'amélioration pour la culture en taillis de peuplier

a) *Valorisation en taillis à courte rotation du progrès génétique dédié à la populiculture classique*

Chez le peuplier, l'installation actuelle de TCR et TtCR repose sur l'utilisation de variétés développées pour un usage en peupleraie classique à très faible densité de plantation et pour la production de grumes (Berthelot and Gavaland 2007). Ces variétés ont été sélectionnées pour leur niveau de productivité élevé associant densité du bois relativement faible, bonne aptitude au déroulage et qualité des fibres pour un usage papetier. Toutefois, il est important de souligner que les objectifs de sélection prioritaires et communs à tous les programmes

développés en Europe et en Amérique du Nord depuis plus de 60 ans concernent une tolérance élevée aux contraintes abiotiques (froid, vent, sècheresse) et biotiques (maladies et insectes ravageurs) qui peuvent diminuer significativement les rendements (FAO 2008).

En France, les premiers travaux visant à évaluer les performances en TCR ou TtCR de cultivars de peupliers présents dans les catalogues nationaux ont été entrepris par l'AFOCEL (devenue FCBA en 2007) dans les années 1970 puis par l'INRA dans les années 1990 (Berthelot and Gavaland 2007). A la même période, des efforts d'évaluation comparables étaient développés en Belgique (Herve and Ceulemans 1996) avec les clones les plus populaires de l'époque, en particulier les hybrides interspécifiques *P. deltoides x P. trichocarpa*, connus pour présenter un très fort effet d'hétérosis pour la croissance. L'apparition de l'épidémie de la rouille foliaire causée par le contournement des résistances complètes portées par l'espèce parente *P. deltoides* a provoqué des dégâts considérables dans les peupleraies conduites en futaie mais également dans tous les TCR utilisant le même matériel végétal (Pinon and Frey 2005). Depuis 2000, les nouvelles expérimentations conduites en Europe évaluent les performances en TCR de deux autres formules hybrides exploitées en futaie (*P.deltoides x P.nigra* ; *P. trichocarpa x P. maximowiczii*) (Bergante et al. 2010; Broeckx et al. 2012; Sylvabiom 2008 - 2012).

Les résultats obtenus sur un à trois cycles de production en TCR ou TtCR montrent que des variétés sélectionnées pour la production de grumes à très faible densité de plantation ne supportent pas forcément des rotations courtes et des récoltes fréquentes (Al Afas et al. 2008a) entrainant des changements de classement des cultivars au cours des cycles successifs. Une forte variabilité individuelle a ainsi été mise en évidence sur l'évolution du taux de mortalité, le nombre de rejets après recépage et le nombre de brins dominants par cépée lors de la récolte (Al Afas et al. 2008a; Broeckx et al. 2012). Rappelons que ces trois caractères qui se révèlent des déterminants majeurs de la productivité en TCR et TtCR n'ont jamais fait l'objet de sélection dans les programmes dont ont été issus les cultivars de peuplier commercialisés. Aucune donnée sur la qualité de la biomasse produite par ces différents clones du commerce n'a été à ce jour publiée.

Si ces résultats ont permis de formuler des recommandations en termes de choix de matériel végétal pour une utilisation spécifique en TCR ou TtCR, ils mettent en évidence l'intérêt de développement de programmes de sélection spécifiques pour une maximisation du progrès génétique et une meilleure valorisation des ressources génétiques. De tels programmes sont actuellement initiés par les différents acteurs de la création variétale du peuplier en Italie, France, Belgique et Allemagne.

b) *Vers de nouveaux idéotypes de peuplier adaptés à la culture en TCR*

La mise sur le marché de clones performants ayant une croissance rapide et un fort rendement grâce à l'amélioration génétique est une priorité pour la réussite des systèmes de culture en taillis (Köhn 2009). Il convient cependant de s'assurer que les critères de sélection utilisés tiennent compte des particularités du système cultural en taillis à courte rotation et de l'évolution possible de ces techniques dans un courant plutôt favorable aux itinéraires à faibles intrants. Pour des utilisations de la biomasse produite ciblées vers la production de biocarburants, il convient de s'intéresser également à une amélioration génétique de la qualité de cette biomasse (Hinchee et al. 2009).

Avant de s'intéresser plus spécifiquement au peuplier, il est intéressant d'analyser les choix réalisés en matière d'amélioration génétique des saules depuis les années 1970, ces programmes visant dès le départ la production de biomasse en TtCR. Les programmes de sélection développés d'abord conjointement puis en parallèle en Suède et au Royaume Uni ont utilisé de larges collections de clones issus d'une centaine d'espèces différentes avant de s'orienter vers la valorisation de clones issus de croisements inter et intraspécifiques. Les sélections phénotypiques ont été basées au départ sur les caractéristiques de la tige (hauteur, diamètre, rectitude), la réaction à la coupe (nombre de rejets de souche, vigueur des rejets), mais aussi sur la résistance aux ravageurs et maladies, nombreux chez les saules, et sur les réponses à des stress environnementaux (Larsson 1998 ; Karp et al. 2011). Ce programme a rapidement mis sur le marché 7 clones élites s'adaptant à de fortes densités de plantation et permettant de doubler la quantité de biomasse produite (Sennerbyforsse and Zsuffa 1995). De nouveaux caractères liés à la physiologie de la canopée mise en place tels que l'efficience d'utilisation de l'eau et l'efficience d'utilisation de l'azote font aujourd'hui l'objet d'analyse de la variabilité génétique chez ces espèces et pourraient être intégrés à moyen terme dans les programmes de sélection (Weih et al. 2011). De même un intérêt croissant est porté à la composition de la biomasse pour différentes technologies de conversion (Tharakan et al. 2003; Karp and Shield 2008; Brereton et al. 2010)

La définition de nouveaux idéotypes de peuplier pour la production de biomasse convertie en biocarburants repose simultanément sur une maximisation de la production de biomasse par unité de temps et sur une optimisation de la qualité de la ressource ligno-cellulosique produite. L'extension attendue des TCR et TtCR de peupliers à l'échelle du territoire national conduit également à ne pas négliger l'adaptabilité à des environnements différents sur le plan climatique et édaphique.

(1) Adaptation aux environnements de production

La phénologie de la croissance reflète un premier niveau d'adaptation aux conditions locales permettant une maximisation de la production de biomasse grâce à un bon calage de la durée de végétation. Un débourrement précoce provoque une mobilisation rapide des ressources, qui va permettre le développement des branches et l'augmentation de la surface foliaire, c'est ainsi qu'on obtient une croissance rapide et une fermeture de la canopée (Sennerbyforsse and Zsuffa 1995; Heilman et al. 1996). Riemenschneideir et al. montrent en 1996 que les clones présentant un débourrement végétatif précoce présentent une grande capacité d'enracinement et une matière sèche racinaire élevée (Riemenschneideir et al. 1996). Un arrêt de croissance et une chute des feuilles trop tardive peuvent se traduire par une mauvaise mise en réserve pouvant compromettre un bon démarrage de la végétation au printemps suivant. Plusieurs études réalisées sur peuplier montrent que des mesures de caractéristiques phénologiques telles que la date de débourrement végétatif mais aussi la date de formation des bourgeons, la coloration des feuilles à l'automne et la chute des feuilles permettent de décrire avec précision les fluctuations de surface foliaire totale de l'arbre au cours de la saison de végétation (Ceulemans et al. 1992; Dunlap and Stettler 1996). Cette information sur l'indice foliaire est centrale dans les modèles de croissance et d'interception de la lumière. Connu pour être sous fort contrôle génétique, ces caractères de phénologie sont également sous important contrôle de facteurs environnementaux : somme de températures journalières en fin d'hiver pour le débourrement végétatif (Dunlap and Stettler 1996; Frewen et al. 2000), photopériode et températures journalières minimales pour l'arrêt de croissance, la coloration et la chute des feuilles (Rohde et al. 2011). Ces caractères jusqu'ici peu pris en compte en sélection pour la populiculture classique présentent un intérêt croissant dans un contexte climatique changeant et peuvent suivre les mêmes orientations de sélection pour différents usages des variétés.

(2) Aptitude à l'enracinement

Peu de choses sont connues sur ce caractère souvent difficile à mesurer sur des grandes collections de génotypes en cours de sélection (Pitre et al. 2010). Aucune corrélation significative n'a été trouvée entre la biomasse racinaire et la biomasse aérienne de 5 clones testés en TCR (Al Afas et al. 2008b). Par contre dans la même étude, la surface de racines fines observée à différentes profondeurs de sol est corrélée positivement avec l'indice foliaire. Toutefois, compte-tenu des densités de plantation élevées utilisées en TCR, il convient de s'assurer d'une bonne aptitude à l'enracinement des clones pour garantir la mise en place du peuplement.

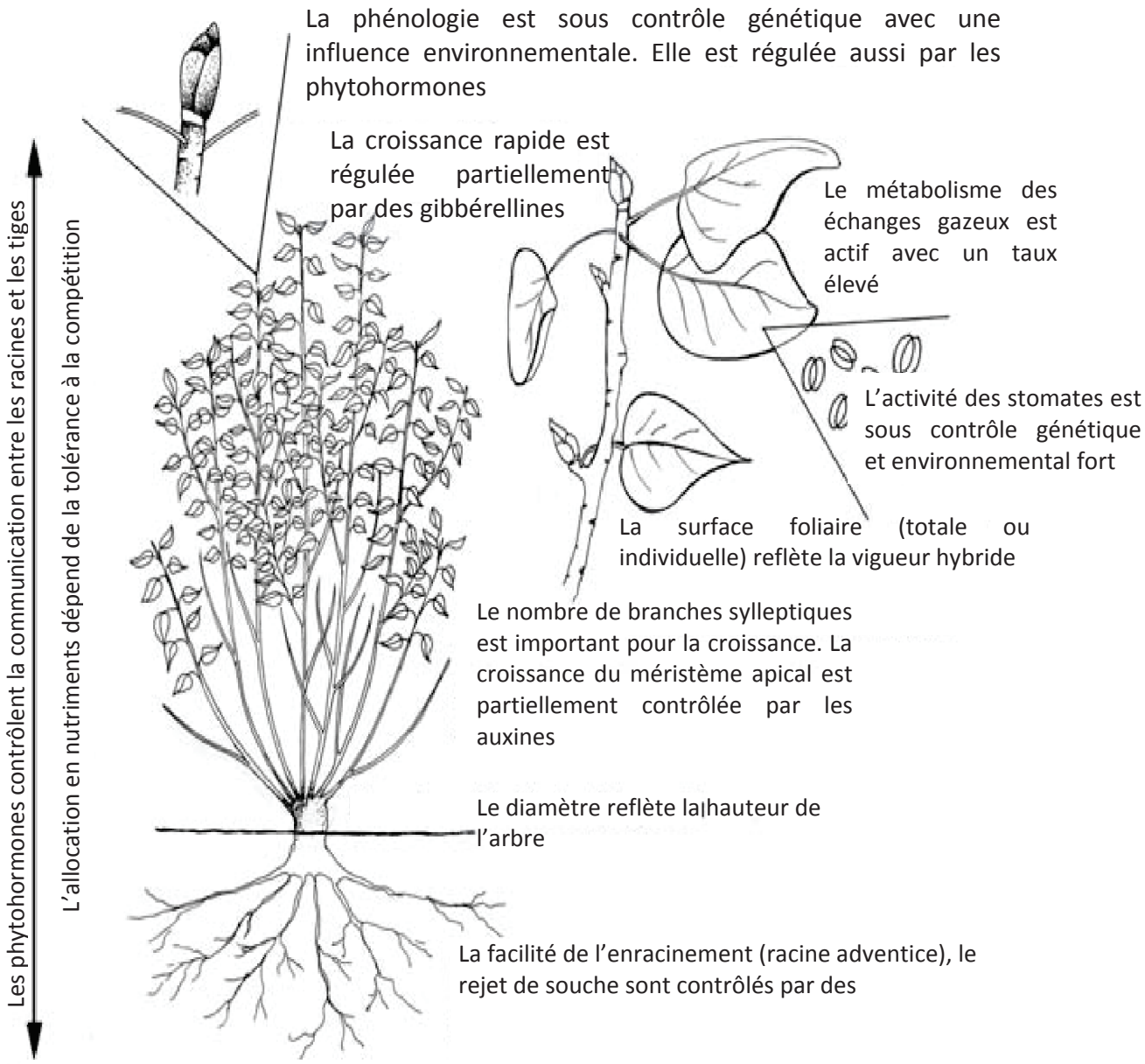


Figure I.3 : Présentation schématique des principales composantes morphologiques et physiologiques affectant la croissance du peuplier conduit en taillis à courte rotation. D'après (Dillen et al. 2010).

(3) Productivité en TCR et TtCR

La production de biomasse est en soi un caractère complexe. La conduite de la culture en TCR rajoute un niveau de complexité en cherchant à maximiser une production sur plusieurs cycles consécutifs de la culture. Au-delà de la croissance initiale lors de l'établissement de la plantation, il s'agit en effet de favoriser une croissance soutenue pour tous les cycles de production issus des mêmes plants.

À l'échelle d'un plant (une cépée en TCR), la croissance dépend de l'interaction entre l'interception lumineuse par la canopée et l'intensité d'assimilation de CO_2 par les feuilles. Plusieurs caractères décrivant soit l'organisation structurale de la canopée, soit le fonctionnement de cette dernière participent à l'élaboration de la productivité (**Figure I.3**).

L'idéotype architectural actuellement proposé pour les TCR ou TtCR présente un nombre de brins élevés après recépage, des feuilles de grande taille, un nombre élevé de branches et plus particulièrement de ramifications sylleptiques (Scarascia-Mugnozza et al. 1999; Berthelot and Gavaland 2007; Dillen et al. 2011). Cet idéotype semble s'éloigner de celui destiné à la populiculture classique. Si une surface foliaire unitaire maximale est visée dans les deux cas, la culture en futaie recherche des arbres faciles à élaguer, présentant peu de branches bien organisées en verticilles annuels. Toutefois, il est important de noter que les traits foliaires et l'architecture de la cépée varient considérablement au sein des espèces pures et des hybrides interspécifiques et il semble encore difficile de statuer sur l'existence d'une seule stratégie efficace pour toutes conditions environnementales et tout système cultural (TCR ou TtCR) (Projet ANR Sylvabiom, appel d'offre en 2008) (Sylvabiom 2008 - 2012).

À l'heure actuelle, plus de 90% des peupleraies sont installées sur des terrains bien alimentés en eau. Le contexte actuel de changement climatique et de gestion raisonnée de la ressource en eau justifie de s'intéresser à de nouveaux caractères tels que l'efficience d'utilisation de l'eau, la tolérance à des stress hydriques modérés voire la résistance à la cavitation en situations de stress hydriques plus marqués (Brignolas et al. 2012a, b). Des travaux combinant évaluations en serre, en pépinière et en peupleraie, en conditions hydriques limitantes et non limitantes, ont révélé l'existence d'une variété individuelle importante au sein d'hybrides commerciaux *P. deltoides x P. nigra* (Monclus et al. 2006; Monclus et al. 2009 ; Fichot et al. 2010). Elle peut même atteindre jusqu'à 40% en terme d'efficience d'utilisation de l'eau. Dans les différentes expériences réalisées, aucune relation défavorable n'a été jusqu'ici mise en évidence entre production de biomasse et efficience d'utilisation de l'eau.

(4) Résistance aux ravageurs et maladies

Dès 1985, Pinon et Schvester définissaient les risques phytosanitaires particuliers encourus par les taillis à courte rotation caractérisés par des densités de plantation élevés, une utilisation de matériel végétal sélectionné et des récoltes mécanisées répétées dans le temps (Pinon and Schvester 1985). Un impact global plus important de certains ravageurs connus a déjà pu être révélé (Pei et al. 2005). Ainsi, la densité des TCR de peuplier semble favoriser la brunissure des feuilles occasionnée par *Marssonina brunnea* dont les spores sont disséminées par rejaillissement. Par ailleurs, la persistance longue sur le végétal d'eau à l'état liquide favorise nettement la germination des spores de champignons pathogènes et peut conduire à des attaques marquées de brunissure des feuilles et de rouilles foliaires occasionnées par *Melampsora-larici-populina* (*Mlp*) ou *Melampsora allii-populina (Map)*. Dans un taillis expérimental en Belgique, l'infection par *Mlp* du cultivar 'Robusta' a réduit fortement sa surface foliaire totale en fin d'été au moment où la surface photosynthétique est maximale dans le peuplement (Broeckx et al. 2012). En peupleraies à large espacement, les pertes de production de biomasse dues aux attaques de rouille foliaire peuvent atteindre certaines années 60% pour des cultivars relativement sensibles (Gastine et al. 2003). Par ailleurs, les souches affaiblies par des attaques répétées de rouilles foliaires, n'émettent que peu des rejets moins vigoureux et sont plus sensibles à des attaques de parasites jugés secondaires généralement (Berthelot and Gavaland 2007). Les conditions de TCR et TtCR étant particulièrement favorables au développement de *Mlp*, il convient d'accorder une priorité encore plus élevé à ce critère dans les sélections à venir.

(5) Qualité du bois

La proportion des trois principaux constituants du bois (cellulose, hémicellulose et lignine) est très déterminante pour les différentes utilisations énergétiques de la biomasse produite (Sannigrahi et al. 2010). Pour des utilisations en combustion et gazéification, des clones avec de fortes teneurs en lignine seront privilégiés (Ohlrogge et al. 2009 ; Dillen et al. 2011). Pour la conversion en biocarburant, la séquestration du carbone dans la biomasse et son partitionnement dans différents compartiments de la plante est très importante pour faciliter l'accès à la cellulose et l'hémicellulose et augmenter ainsi le rendement en saccharification (Ragauskas et al. 2006). Idéalement, les propriétés recherchées pour un rendement élevé en saccharification sont des teneurs élevées en cellulose et en hémicellulose avec *a contrario* une faible teneur en lignines (Sannigrahi et al. 2010; Novaes et al. 2010).

La lignine qui est un constituant majeur du bois représente entre 21 et 29% de la matière sèche chez le peuplier (Sannigrahi et al. 2010). Le terme lignine est en fait générique et désigne

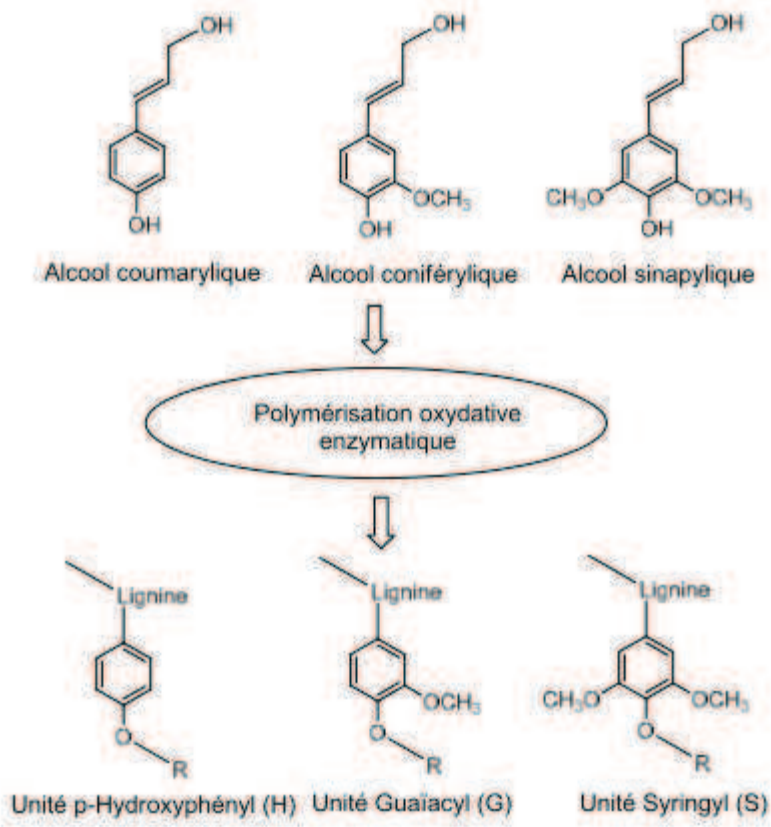


Figure I.4 : Représentation des 3 molécules d'hydroxy-cinnamyl alcools à la base de la structure moléculaire des polymères de lignine.

un ensemble de biopolymères aromatiques obtenu par le couplage oxydatif de 4-hydroxyphenylpropanoïdes (Ralph et al. 2004). Ces polymères sont déposés dans la paroi secondaire des cellules végétales et lui confèrent rigidité et imperméabilité. La structure des lignines peut être complexe (Boerjan et al. 2003), cependant, les unités de base du polymère sont les hydroxy-cinnamyl alcools aussi appelés monolignols. Il en existe 3 principaux qui sont le coniferyl alcool, le synapil alcool et le p-coumaryl alcool qui donnent respectivement les 3 phenylpropanoïdes guaiacyl (G), syringyl (S) et p-hydroxyphenyl (H) une fois polymérisés (**Figure I.4**). Si la lignine des angiospermes est principalement composée des sous unités S et G, celle des gymnospermes est composée en majorité d'unités G avec une faible proportion de H (Campbell and Sederoff 1996). Différentes études démontrent l'importance du rapport S/G dans le processus de délignification (Chang and Sarkanen 1973; Chiang and Funaoka 1988). Le ratio S/G a une influence sur l'impact négatif d'un taux de lignines élevé sur le rendement en saccharification, en effet, lorsqu'il atteint des valeurs élevées la corrélation négative observée diminue voire disparait (Davison et al. 2006; Studer et al. 2011).

Dans la biomasse végétale, ce sont la cellulose et l'hémicellulose qui constituent la matière première pour la conversion en bioéthanol (Ragauskas et al. 2006; Rubin 2008). Ces composés sont enchevêtrés dans les lignines au sein de la paroi secondaire des cellules végétales et sont donc difficilement accessibles par les enzymes de la saccharification. Au cours du processus industriel, la biomasse ligno-cellulosique subit des prétraitements chimiques (attaque acide, basique, explosion à la vapeur) et/ou physiques (broyages, températures élevées) préalables, afin de libérer la cellulose et l'hémicellulose pour les rendre accessibles aux enzymes de la saccharification et ainsi améliorer le rendement en sucres fermentescibles. Ces sucres sont majoritairement du glucose pour les C6, et du xylose pour les C5 (Sannigrahi et al. 2010).

Dans l'objectif d'améliorer le rendement de la saccharification, des recherches sont engagées dans plusieurs voies, allant de l'optimisation de la qualité de la ressource, aux ajustements des conditions de prétraitement de la biomasse (Wang et al. 2012), ou de l'hydrolyse enzymatique (Navarro et al. 2010; Jagtap et al. 2012), et de leurs potentielles interactions (Studer et al. 2011). En ce qui concerne l'optimisation de la qualité de la ressource ligno-cellulosique, les attentes sont principalement dirigées vers les propriétés chimiques de la biomasse, à savoir les taux de cellulose, d'hémicellulose et de lignines, mais aussi la qualité des lignines évaluée par le ratio entre les monomères syringyl et guaiacyl (ratio S/G), ainsi que la qualité de la cellulose évaluée par son degré de cristallinité (Sannigrahi et al. 2010 ; Ragauskas et al. 2006 ; Studer et al. 2011 ; Chen and Dixon 2007). Un faible degré de cristallinité est généralement recherché pour la saccharification (Puri 1984; Sannigrahi et al. 2010). Des études sur le bois de tension,

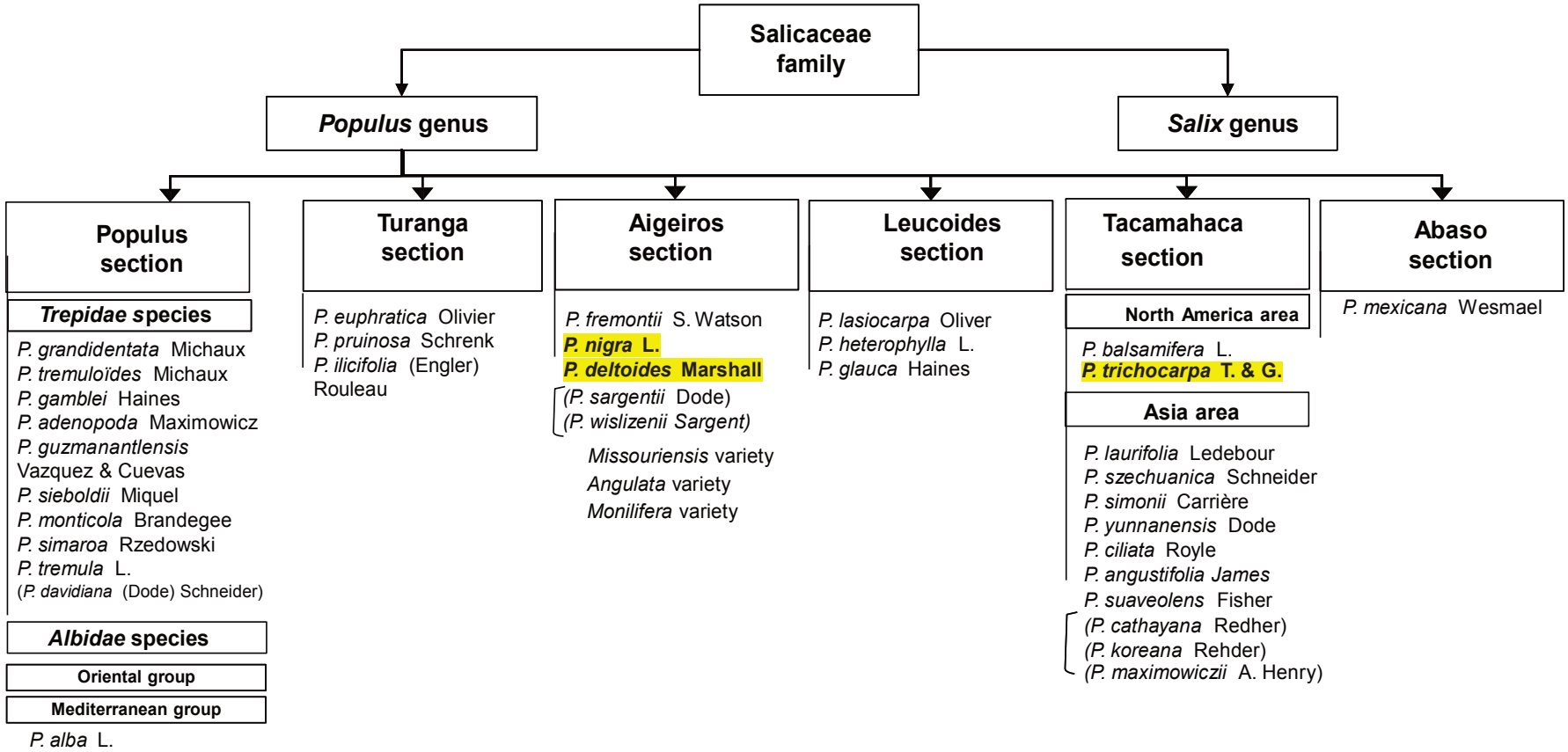


Figure I.5 : Subdivision du genre *Populus* en plusieurs sections. *P. deltoides*, *P. trichocarpa* et *P. nigra* sont les 3 principales espèces utilisées en amélioration génétique. D'après (Dillen et al. 2011)

caractérisé par une très haute teneur en cellulose cristalline, ont toutefois révélé des rendements accrus en saccharification (Brereton et al. 2011 ; Foston et al. 2011).

C. Ressources génétiques du peuplier

Les 29 espèces reconnues au sein du genre *Populus* sont taxonomiquement réparties en six sections (**Figure I.5**). Ces espèces sont en très grande majorité diploïdes (2n = 38). Leurs aires naturelles de distribution, souvent vastes, sont toutes situées dans l'hémisphère Nord. Les populations naturelles s'organisent le plus souvent le long des ripisylves et dans les zones humides (Dickmann and Kuzovkina 2008). Toutefois certaines espèces telles que les trembles (*P.tremuloides, P. tremula*) forment des peuplements forestiers assez vastes. Aucun peuplier ne pousse à l'état spontané dans l'hémisphère Sud mais il peut être cultivé à des latitudes équivalentes de celles sous lesquelles il est cultivé dans l'hémisphère Nord. Les peupliers sont peu longévifs par rapport à d'autres arbres, néanmoins, certains arbres peuvent atteindre 200 à 300 ans, et les systèmes racinaires des peupliers trembles (section *Populus*) peuvent persister des milliers d'années pour de multiples générations de tiges (Dickmann and Kuzovkina 2008).

1. Ressources cultivées

Plusieurs de ces espèces sont importantes en sylviculture, et sont largement plantées dans le monde entier y compris dans l'hémisphère sud (Dickmann and Kuzovkina 2008; Zamudio and Yanez 2010). En 2008, la commission internationale du peuplier (CIP) a enregistré un total de 5,3 millions d'hectares de plantations de peupliers dans le monde. C'est la Chine qui vient en tête avec 4,3 millions ha, vient après la France avec 236 000 ha, la Turquie avec 125 000 ha, l'Italie avec 118 500 ha et l'Allemagne avec 100 000 ha (FAO 2008). Toutefois dans tous ces pays, la part des plantations spécifiquement dédiée à des fins énergétiques est très faible: seulement 0,9% de leur production totale de bois (bois de peuplier et de saule) a été utilisée comme bois de feu ou en tant que ressource de bioénergie (FAO 2008). Les TCR et TtCR à vocation énergétique sont en grande partie expérimentaux. Au nord de l'Italie, pays d'Europe le plus dynamique sur l'espèce, plus de 4000 ha de TCR et TtCR de peuplier sont en test pour estimer la rentabilité de ce genre de culture énergétique, les effets de l'environnement et des clones et d'optimiser les régimes de culture (densité, rotation, récolte etc.) (Spinelli et al. 2008 ; Bergante et al. 2010).

Depuis plus d'une centaine d'années, les ressources génétiques cultivées sont à plus de 95% des clones issus d'une sélection phénotypique intraspécifique ou des clones issus d'hybridations interspécifiques. L'hybridation interspécifique se produit naturellement entre espèces sympatriques d'une même section mais est aussi possible par croisements contrôlés entre

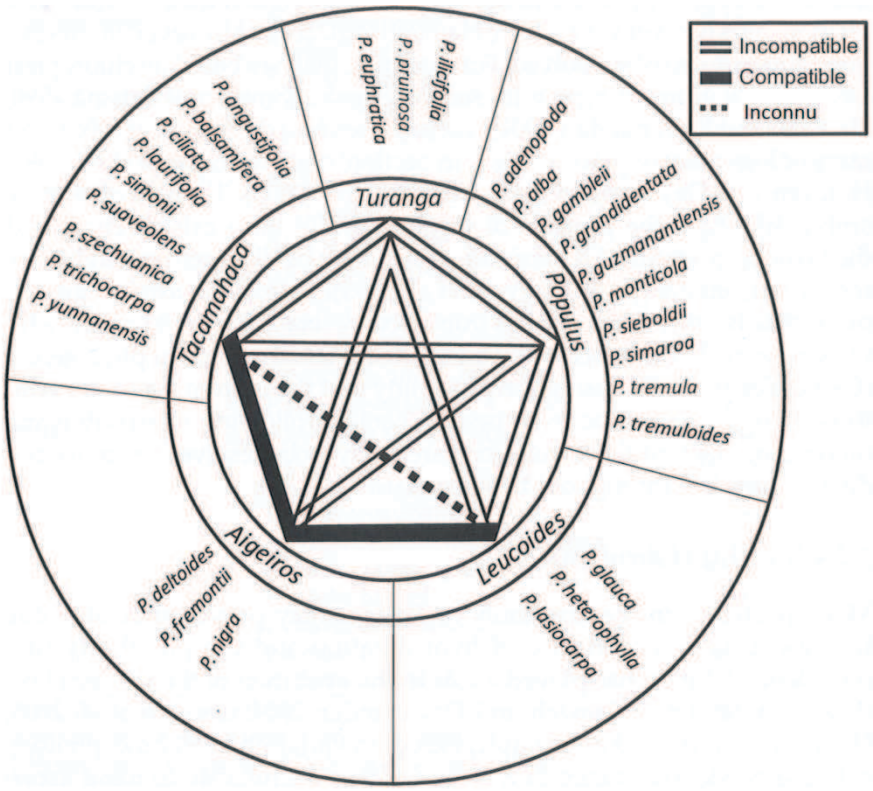


Figure I.6 : Compatibilité de croisements entre espèces et sections du genre *Populus*.
D'après Willing and Pryor (1976) repris par (DiFazio et al. 2011)

certaines espèces d'une même section ou de sections différentes (**Figure I.6**). L'hybridation peut parfois être unidirectionnelle. Ainsi l'hybridation entre *P. nigra* et *P. deltoides* n'est possible que dans le sens *P. deltoides* femelle x *P. nigra* mâle.

Le registre international des cultivars de peuplier commercialisés dans le monde entier, géré par la CIP sous l'égide de la FAO comportait 332 accessions en septembre 2012 (http://www.populus.it/ipccloni.php). Celles-ci représentaient:

- 128 cultivars issus de 7 espèces pures : 57 cultivars de *P. deltoides*, 40 cultivars de *P. nigra*, 19 cultivars de *P. trichocarpa*, 13 cultivars de *P. alba*, 6 cultivars de *P. tomentosa*, 2 cultivars de *P. tremula* et un cultivar de *P. tremuloides*.
- 204 cultivars issus d'hybridation interspécifique dont 147 cultivars *P. x canadensis* issus d'hybridation entre *P. deltoides* et *P. nigra*, 17 cultivars *P. x canescens* issus d'hybridation entre *P. tremula* et *P. alba*, 13 cultivars *P. x interamericana* issus de croisements entre *P. deltoides* et *P. trichocarpa*, 6 cultivars issus d'hybridation entre *P. nigra* et *P. maximowiczii* et 21 cultivars issus d'hybridations plus complexes entre les 7 espèces pures citées précédemment.

Deux remarques très importantes sont à signaler :

- Les ressources génétiques offertes par ce registre international sont actuellement sous-exploitées. Ainsi les statistiques de ventes de cultivars pour l'année 2007 dans quatre pays européens (source rapport CIP-FAO 2008) montraient que respectivement 9, 7, 16 et 5 clones représentaient à eux seuls 90% des ventes de plants en France, Italie, Belgique et Espagne respectivement. Parmi ces clones très populaires, le clone I-214, sélectionné en 1929, représentait la même année 24%, 77% et 70% des ventes en France, Italie et Espagne respectivement.
- Plusieurs de ces clones sont fortement apparentés. À titre d'exemple, les 13 cultivars *P. x interamericana* comportent 4 cultivars plein-frères et 8 cultivars demi-frères.

Chez le peuplier, le recours à l'hybridation interspécifique a 3 buts majeurs :
(i) combiner les traits favorables présents chez des espèces différentes, (ii) obtenir une vigueur hybride qui peut être définie comme la supériorité des hybrides sur le meilleur des deux parents, (iii) favoriser l'homéostasie qui peut être définie comme la stabilité des phénotypes dans différents environnements (Stettler et al. 1996). Le succès connu depuis près de 60 ans par les hybrides interspécifiques s'explique essentiellement par les deux premiers points. Ainsi, les hybrides *P. x canadensis* combinent le bon enracinement et un bon niveau de résistance à la rouille foliaire du parent *P. nigra* et la vigueur juvénile élevée du parent *P. deltoides*. En matière

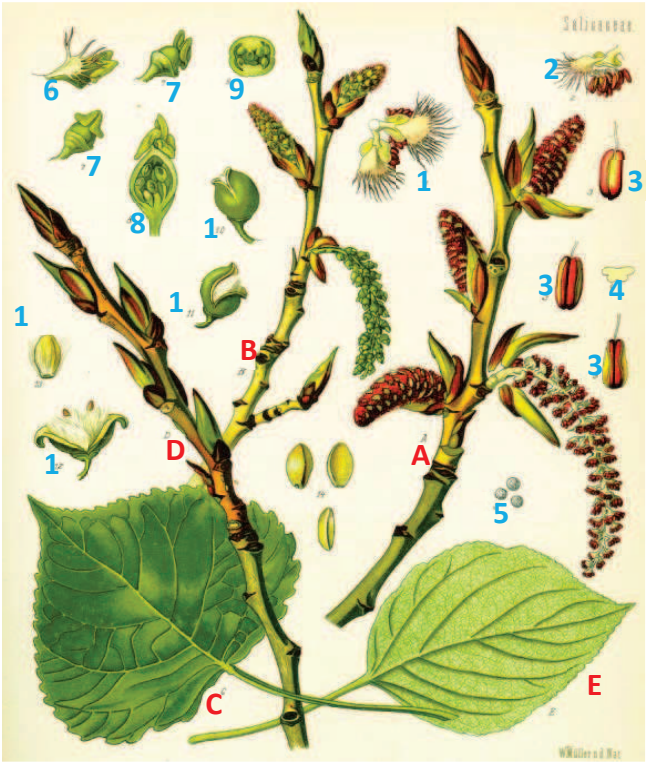


Figure I.7 : Caractéristiques botaniques du peuplier noir (Köhler's Medizinal-Pflanzen 1897).

A. Rameau floral mâle, B. Rameau floral femelle, C. Feuille de peuplier noir, D, Rameau avec bourgeons, E. Feuille de peuplier d'Italie, 1. et 2. Fleur mâle avec sa bractée, 3. Étamine, 4. Étamine en coupe transversale, 5. Grains de pollen, 6. Fleur femelle avec sa bractée, 7. Idem sans la bractée, 8. Ovaires en coupe longitudinale, 9. Idem en coupe transversale, 10., 11. et 12. Capsule à différents temps de maturation, 13. Graine.

Figure I.8 : Aire de répartition naturelle du peuplier noir (Dickmann and Kuzovkina 2008)

de vigueur hybride pour la productivité, plusieurs études s'intéressant à différentes combinaisons parentales évaluées dans différents environnements s'accordent à souligner des gains de croissance variant entre +40 et +70% entre familles (Dillen et al. 2009) et (GIS-Peuplier/INRA, données non publiées) et pouvant atteindre +300% au niveau individuel pour un hybride F1 *P. trichocarpa* x *P. deltoides* (Stettler et al. 1988) justifiant l'intérêt pour l'hybridation interspécifique lorsqu'une diffusion du progrès génétique par voie clonale est possible.

L'intérêt marqué en Europe pour les hybrides *P. x canadensis* entre *P. deltoides* et *P. nigra* conduit les généticiens et sélectionneurs à s'intéresser plus particulièrement aux ressources génétiques disponibles au sein de l'espèce autochtone *P. nigra*.

2. Exploration des ressources naturelles chez une espèce endémique *Populus nigra* (*P. nigra*)

a) *Présentation de l'espèce Populus nigra*

Le peuplier noir est une espèce longévive, dioïque, fonctionnant en métapopulation et atteignant la maturité sexuelle à partir de 5-6 ans (Zsuffa 1974) (**Figure I.7**). Son aire de distribution naturelle s'étend de l'Europe de l'Ouest à l'extrémité Ouest de la Chine en passant par une étroite frange Nord-Africaine (**Figure I.8**, (Dickmann and Kuzovkina 2008)). *P. nigra* est l'espèce dominante de la ripisylve et contribue fortement à la biodiversité remarquable dans cet écosystème. *P. nigra* est une espèce dominante et emblématique des rives des cours d'eau français et partage souvent son habitat avec deux autres espèces autochtones de peuplier, le peuplier blanc *P. alba* et le peuplier tremble *P. tremula*. La tolérance du peuplier noir aux faibles températures extrêmes lui a permis de s'établir sur des terrains couvrant une gamme altitudinale importante. Le peuplier noir est reconnu pour son rôle épurateur important dans l'amélioration de la qualité des eaux notamment grâce à un système racinaire très développé qui joue le rôle de filtre naturel en piégeant certains polluants comme les phosphates et les nitrates d'origine agricole ou urbaine, en plus de ce rôle épurateur, le système racinaire du peuplier noir est un facteur déterminant dans la cohésion et la fixation des matériaux alluvionnaires (Ruffinoni and Gazelle 1997).

Le peuplier noir est une espèce anémogame, le pollen est produit en grande quantité puis transporté par le vent jusqu'aux fleurs femelles au cours des mois de mars et avril. Les fruits se présentent comme des capsules ovoïdes à deux valves qui libèrent leurs graines entourées d'un coton hydrophobe de fin avril à mi-juillet (Zsuffa 1974). Six à huit semaines après fécondation débute la dissémination des graines. Le succès de la germination est lié à la concordance temporelle entre la dispersion des graines et l'existence de sites favorables mis à nus après des

crues. La survie des jeunes semis est étroitement dépendante de l'adéquation entre croissance racinaire et fluctuation du niveau de la nappe. De nombreux auteurs s'accordent à penser que la survie des semis la première année constitue l'une des étapes les plus critiques pour le succès du maintien de la population (Chamaillard 2011).

Étant inféodé à la ripisylve, le peuplier noir subit les mêmes menaces que son habitat : altération, fragmentation et réduction de surface par l'agriculture et les activités humaines s'y rapportant (Lefevre et al. 1998; Hughes and Rood 2003).

b) *Diversité génétique neutre et conservation des ressources naturelles*

Depuis la fin des années 90 et le développement de marqueurs moléculaires, les études traitant de la diversité génétique neutre du peuplier noir se sont multipliées. Elles sont venues compléter des travaux sur la diversité génétique adaptative initiés par le projet européen EUROPOP (1997-2001) qui s'est intéressé aux populations naturelles au sein de divers systèmes fluviaux européens (Van Dam and Bordacs 2002). Les populations de peuplier noir ayant la taille efficace la plus élevée se trouvent dans le Sud de l'Europe (Storme et al. 2004). En effet, en Belgique, en Allemagne et aux Pays-Bas, et plus généralement dans le Nord-Ouest de l'Europe (Grande-Bretagne incluse), le peuplier noir n'est plus présent qu'à l'état de relique avec des peuplements de faible taille, dispersés et souvent constitués de copies végétatives maintenues par l'Homme (Koskela et al. 2004 ; Smulders et al. 2008a). Le calcul des indices de diversité calculé à partir de 7 marqueurs microsatellites ou SSR (Short sequence repeat) a mis en évidence une diversité élevée au sein des populations (Ho: 0,70 – 0,83), exceptée pour la population du Royaume-Uni qui était constituée à 97% de génotypes répliqués (Smulders et al. 2008b). Les populations échantillonnées le long d'une même rivière sont proches génétiquement (Fst: 0,002 – 0,037) et il existe une différenciation faible entre rivières (Fst = 0.081). Dans une autre étude réalisée le long de la Drôme où il existe des zones favorables à la régénération naturelle en amont alors que la ripisylve tend à être limitée en aval par l'agriculture, Imbert et Lefèvre (2003) ont révélé avec 6 marqueurs SSR une diversité génétique globale élevée et une différenciation entre populations traduisant un isolement par la distance (Imbert and Lefevre 2003). En comparaison avec les populations naturelles présentes sur différents réseaux fluviaux européens, la diversité génétique du peuplier noir reste élevée en France en moyenne et présente une légère différenciation entre les réseaux déconnectés (Lefèvre et al. 2002; Storme et al. 2004; Smulders et al. 2008b).

Tableau I.3 : Niveaux de variabilité génétique et d'héritabilités estimés chez *P. nigra* pour différents caractères d'intérêt. Héritabilité au sens strict $h^2 = V_A/V_P$ où V_A est la variance génétique additive et V_P la variance phénotypique. Héritabilité au sens large $H^2 = V_G/V_P$ où V_G est la variance génétique totale et V_P la variance phénotypique. La variabilité géographique concerne des populations naturelles dans différents bassins fluviaux ; la variabilité familiale concerne des familles de demi-frères ou de plein-frères obtenues par pollinisation libre ou par croisement contrôlé ; la variabilité clonale correspond à la variabilité entre clones et peut cumuler variabilité entre familles et variabilité entre populations.

Caractère	Variabilité génétique			Héritabilités		Références
	Géographique	Familiale	Clonale	Sens strict h^2	Sens large H^2	
Biomasse						
Croissance en hauteur	modérée	élevée	élevée	0,32 - 0,42	0,31 - 0,67	(Teissier du Cros 1977) (Nepveu et al. 1978) (Pichot and Teissier du Cros 1988) (Legionnet et al. 1999) (Isik and Toplu 2004) (Guet 2011)
Croissance en circonférence	modérée	élevée	élevée	0,15 - 0,55	0,34 - 0,82	(Teissier du Cros 1977) (Nepveu et al. 1978) (Isik and Toplu 2004) (Guet 2011)
Biomasse aérienne	modérée	élevée	élevée	-	0,28 - 0,77	(Chamaillard et al. 2011)
Phénologie						
Précocité du débourrement	élevée	élevée	élevée	0,61 - 0,70	0,72 - 0,86	(Teissier du Cros 1977) (Pichot and Teissier du Cros 1988)
Arrêt de croissance	élevée	élevée	élevée	0,32 - 0,49	0,46 - 0,76	(Teissier du Cros 1977) (Rohde et al. 2011) (Fabbrini et al. 2012)
Résistance aux ravageurs	faible	modérée	élevée	-	0,22 - 0,59	(Teissier du Cros 1977) (Pichot and Teissier du Cros 1988) (Legionnet et al. 1999)
Qualité du bois						
Infradensité du bois	-	-	élevée	-	0,61 - 0,69	(Nepveu et al. 1978)

Les pools génétiques *P. nigra* utilisés dans les programmes d'amélioration génétique européens de peuplier correspondent le plus souvent à des sélections d'individus remarquables dans ces populations naturelles. Leurs tailles varient de quelques dizaines d'individus dans les programmes conduits en Belgique et aux Pays-Bas, à quelques centaines d'individus dans les programmes français et italiens.

Des programmes de conservation des ressources génétiques sont coordonnés à l'échelle nationale par le Ministère de l'Agriculture et à l'échelle européenne par le réseau EUFORGEN (European Forest Genetic Resources Program) depuis 1993. Deux stratégies de conservation complémentaires sont adoptées. D'une part la création de collections *ex-situ* rassemblant des clones échantillonnés dans les ripisylves où la régénération naturelle ne peut plus se faire, d'autre part la gestion *in-situ* de populations naturelles de référence au sein de réserves naturelles (Lefevre et al. 1998). En France, la collection nationale du peuplier noir comprend actuellement 350 individus échantillonnés dans différentes populations et est conservée sous forme de parc à pieds mères à la pépinière forestière d'État de Guemené Penfao. La préparation de chartes de gestion d'unités conservatoires concerne actuellement 14 populations incluses dans des réserves naturelles et devrait être élargie à de nouvelles populations suite à la caractérisation en cours de la diversité génétique de 46 populations naturelles (M. Villar, comm. pers.)

c) *Variabilité génétique pour des caractères adaptatifs et de production*

Plusieurs études se sont attachées à estimer le niveau de variabilité génétique (clonale) pour différents caractères adaptatifs et de production de bois. Elles concernent le plus souvent des collections d'individus non apparentés, issus de repérage en populations naturelles. Quelques plans de croisement intraspécifique ou des comparaisons parents-descendants ont toutefois permis de comparer les parts de variance phénotypique expliquées par les effets génétiques additifs (héritabilité au sens strict h^2) ou par les effets génétiques totaux (héritabilité au sens large H^2). Tous ces travaux révèlent des niveaux de variabilité génétique individuelle plutôt élevés pour la phénologie de croissance et de floraison (Teissier du Cros 1977; Pichot and Teissier du Cros 1988; Rohde et al. 2011), la croissance juvénile (Teissier du Cros 1977; Nepveu et al. 1978; Pichot and Teissier du Cros 1988; Isik and Toplu 2004), la résistance à la rouille foliaire (Teissier du Cros 1977; Pichot and Teissier du Cros 1988; Legionnet et al. 1999), la densité du bois (Teissier du Cros 1977; Nepveu et al. 1978; Isik and Toplu 2004) (**Tableau I.3**). Les héritabilités au sens large sont de modérées à élevées pour les différents caractères alors que les héritabilités au sens strict sont plus variables. Quelques études permettent de connaitre le niveau

de différenciation génétique entre populations naturelles pour des caractères adaptatifs ou fonctionnels. L'expression de la variabilité génétique de la date d'arrêt de croissance et de la formation du bourgeon terminal a ainsi été récemment étudiée au sein d'une large collection européenne de populations naturelles de peuplier noir (Rohde et al. 2011). Cette étude a permis de révéler une différenciation latitudinale des populations : les populations les plus nordiques arrêtaient leur croissance plus tôt que les populations méridionales suggérant une adaptation aux conditions locales de photopériode et de température. Cette différenciation impacte significativement la croissance des populations faisant l'objet d'un transfert latitudinal marqué dans les expérimentations analysées (Guet 2011). Une différenciation entre trois populations naturelles de Loire, de la Drôme et des Pyrénées a également été mise en évidence pour l'efficience d'utilisation de l'eau (Isik and Toplu 2004 ; Chamaillard 2011; Guet 2011) et pourrait également révéler d'autres adaptations locales à des contraintes hydriques plutôt limitantes. Si une forte variabilité individuelle pour la résistance à la rouille foliaire à *Mlp* a été mise en évidence, aucune différenciation entre populations n'a été mise en évidence à l'échelle du bassin fluvial de la Loire (Legionnet et al. 1999).

L'importante variabilité génétique pour plusieurs caractères cibles de la sélection en populiculture traditionnelle explique l'utilisation croissante de ces ressources génétiques dans les programmes de création variétale. Les stratégies d'amélioration à court terme reposent essentiellement sur la valorisation de la variabilité génétique individuelle en croisement interspécifique *P. deltoides* x *P. nigra* du fait des possibilités de clonage à grande échelle de matériel élite identifié (Giordano 1973; Gergacz 1978; Krzan 1981; Schoeneman 1986; Slycken et al. 2002). La facilité de cette multiplication végétative chez le peuplier et surtout chez les espèces appartenant aux sections *Tacamahaca*, *Aigeiros* et chez leurs hybrides est le secret du succès de l'amélioration du peuplier à travers le monde (Riemenschneideir et al. 2001 ; Dickmann and Kuzovkina 2008). En France et en Italie, les stratégies d'amélioration développées à plus long terme correspondent à des schémas de sélection récurrents réciproques consistant en une ou plusieurs générations d'hybridation intraspécifique valorisant la diversité génétique des pools parentaux *P. deltoides* et *P. nigra* et leur complémentarité en croisement interspécifique avant de finaliser la sélection individuelle en hybridation interspécifique (Pichot and Teissier du Cros 1988; Bisoffi and Gullberg 1996).

D. Un programme de sélection efficace à court terme

1. Des contraintes à lever

Pour répondre rapidement aux demandes spécifiques de matériel élite de peuplier pour la production de biomasse de qualité en TCR et TtCR, les programmes d'amélioration cherchent encore à associer (1) l'hybridation interspécifique *P. deltoides x P. nigra* pour la complémentation de caractères et la vigueur hybride observée sur la croissance, (2) une valorisation de la variabilité génétique présente au sein de l'espèce *P. nigra,* (3) l'importante variabilité individuelle et la multiplication végétative. Deux facteurs limitants principaux freinent actuellement l'identification et la diffusion de clones très performants issus d'une très forte sélection au sein d'un très grand nombre de candidats obtenus par croisements contrôlés. Il s'agit d'une part des capacités d'évaluation phénotypique des différents caractères cibles et d'autre part de la possibilité d'une sélection précoce dès le premier cycle de production des taillis.

L'évaluation phénotypique d'un très grand nombre de candidats pour des caractères de production nécessite des surfaces de plantation importantes et mobilise des moyens humains importants pour l'installation des dispositifs expérimentaux et la réalisation des observations phénotypiques en particulier lors des opérations de récolte de biomasse (Riemenschneideir et al. 2001). L'évaluation phénotypique de résistances aux ravageurs nécessite elle aussi des moyens importants lorsque l'on veut contrôler niveau et homogénéité de l'inoculation et variabilité génétique des populations de ravageurs. L'évaluation phénotypique des composants chimiques du bois repose actuellement sur des méthodes de référence lourdes et coûteuses (cf. chapitre V) et ne peut être envisagée en routine que grâce au développement de méthodes d'évaluation indirecte à haut débit.

Les possibilités de sélection précoce pour les différents caractères cibles sont peu connues sauf pour la résistance à *Mlp* qui peut être réalisée très tôt en conditions naturelles d'infection suffisante et en tests artificiels au laboratoire. En ce qui concerne la croissance, on peut toutefois penser qu'une sélection en plusieurs étapes consistant à éliminer en premier cycle de production plus de 50% des individus les moins performants laisse des possibilités de gain génétique important pour la productivité sur plusieurs rotations. Des expérimentations spécifiques devront être mises en place pour étudier la stabilité de la valeur qualitative de la biomasse produite dans les différentes phases de développement et récolte des taillis.

C'est la contrainte imposée par des corrélations juvéniles-adultes faibles qui a été le moteur du développement de marqueurs moléculaires chez les arbres forestiers en vue du développement d'une sélection assistée par marqueurs (SAM) (Grattapaglia and Sederoff 1994). Les différents

travaux conduits pendant une dizaine d'années sur l'architecture génétique de caractères complexes chez différentes espèces ont rapidement révélé des difficultés de mise en application de cette SAM. D'une part, le nombre de régions génomiques d'intérêt identifiées pour un même caractère chez une espèce donnée s'avère élevé lorsqu'on cumule les études réalisées dans des fonds génétiques différents, à des âges différents, dans des environnements différents. Ainsi pour le peuplier, plus de 600 régions génomiques ou quantitative trait loci (QTL) ont été identifiés pour la croissance (Bradshaw and Stettler 1995; Wu et al. 1998; Wullschleger et al. 2005; Street et al. 2006; Zhang et al. 2006; Rae et al. 2007; Rae et al. 2008; Dillen et al. 2009; Novaes et al. 2009; Rae et al. 2009; Marron et al. 2010; Street et al. 2011). La part de variation expliquée par ces QTL est souvent relativement faible expliquant jusqu'à 36%. Peu de caractères à contrôle oligogénique ou à loci à forts effets, constituant les cas plus favorables au développement de SAM ont été identifiés. Les quelques cas observés chez le peuplier concernent la résistance à la rouille foliaire à *Mlp* où les résistances majeures identifiées se sont malheureusement révélées parfois peu durables (Cervera et al. 1996; Jorge et al. 2005; Dowkiw et al. 2010; Bresson et al. 2011). Des approches combinant la détection de QTL et la cartographie de gènes candidats ont été menées sur plusieurs modèles forestiers en particulier pour plusieurs composants chimiques du bois dont les voies de biosynthèse étaient partiellement connues (Neale et al. 2002; Freeman et al. 2009; Gion et al. 2011). Sur Eucalyptus (Thumma et al. 2010) et sur peuplier (Novaes et al. 2009), des régions génomiques à effet majeur pour des propriétés du bois, communes à différentes espèces et contenant des gènes connus pour leur rôle dans des voies de biosynthèse ont été identifiées. Ces résultats cumulant les informations recueillies par plusieurs méthodes d'exploration des génomes et de leur expression redonnent espoir quant à l'utilisation d'une information moléculaire en sélection. La précision de la prédiction offerte par les marqueurs moléculaires dépend beaucoup de l'étendue du déséquilibre de liaison qui doit être élevé entre les marqueurs à utiliser et les locus impliqués dans le caractère à sélectionner. Il est donc crucial de disposer d'un grand nombre de marqueurs polymorphes afin d'identifier ceux qui sont le mieux associés aux allèles d'intérêt. Le développement de ces marqueurs moléculaires reste complexe et passe par différentes phases d'investigations et par la valorisation de différentes ressources génomiques disponibles sur les espèces étudiées.

2. Ressources génomiques disponibles chez le peuplier

a) *Une séquence de référence chez P. trichocarpa*

L'intérêt de disposer d'une séquence de référence complète et précise constitue le seul moyen de disposer d'un catalogue complet des gènes en les annotant bio-informatiquement et en

leur assignant des fonctions par comparaison à d'autres espèces. L'ordre précis des gènes sur la séquence est une information essentielle aux sélectionneurs pour la sélection assistée par marqueurs et le clonage positionnel. La connaissance de ces gènes est d'une grande utilité pour les études d'expression (transcriptomique), les puces peuvent ainsi intégrer tous les modèles de gènes et une analyse plus approfondie des ARM_m permet d'identifier les réseaux de gènes impliqués dans un tissu donné ou répondant à un stress donné.

C'est l'individu Nisqually-1, femelle de l'espèce *P. trichocarpa*, récolté sur les rives du fleuve Nisqually (130 km de long) à l'Ouest de l'État de Washington au Nord-Ouest des États-Unis, qui a été séquencé. Ce séquençage a été effectué en whole genome shotgun (WGS) à hauteur de 7,5 X comme profondeur de séquençage (Tuskan et al. 2006) par le Joint Genome Institute (JGI). Le JGI a effectué le premier assemblage du génome (v1) aboutissant à une séquence de 486 Mb. Les 19 chromosomes représentaient 308 Mb au total et 21993 scaffolds non assignés représentaient 178 Mb. Cette première version d'assemblage a donné naissance à deux versions d'annotation la v1.0 et la v1.1 qui abritaient 45 555 modèles de gènes

Le séquençage du génome a révélé un événement de duplication du génome entier qui se caractérise par 8 000 gènes dupliqués (**Figure I.9**). La comparaison de ces gènes avec les EST (Expressed Sequence Tag) des espèces du genre *Salix* a montré que cette duplication était antérieure à la séparation entre les deux genres, il y a 60 à 65 millions d'années. Des événements de réarrangement après cette duplication ont été observés, qui seraient antérieurs à la séparation entre les différentes espèces du genre *Populus*; en effet, les cartes génétiques de différentes espèces de peuplier ont montré une bonne colinéarité.

La collaboration entre le JGI et le centre de la génomique intégrative a donné naissance à Phytozome dans le but d'effectuer des recherches en génomique comparative entres espèces végétales (Goodstein et al. 2012) (http://www.phytozome.org/). Phytozome a réalisé deux assemblages majeurs du génome, une version 2 qui était disponible début 2010 et une version 3 qui a été réalisée en août 2012. Dans cette thèse nous avons essentiellement utilisé les versions v1 et v2, la v3 étant disponible à un stade tardif de la thèse. La version 2 abritait 2 518 scaffolds (les 19 chromosomes sont appelés aussi scaffolds) représentant 417 Mb. Les 19 premiers scaffolds représentaient 378 Mb et les 2 499 scaffolds non assignés pesaient 39 Mb. L'annotation (v2.0) du second assemblage a révélé 41 377 modèles de gènes avec 45 778 transcrits.

Une version v2.2 a été établie, elle ne diffère pas beaucoup de la v2.0, elle a permis de révéler 40 668 modèles de gènes et 45033 transcrits. La version d'assemblage v3 représente aujourd'hui 1446 scaffolds ayant une taille totale de 422,9 Mb. Les 19 chromosomes sont constitués de 388 Mb. Parmi les 1446 scaffolds, 181 dépassent 50 kb en taille et représentent ainsi 97,3% du

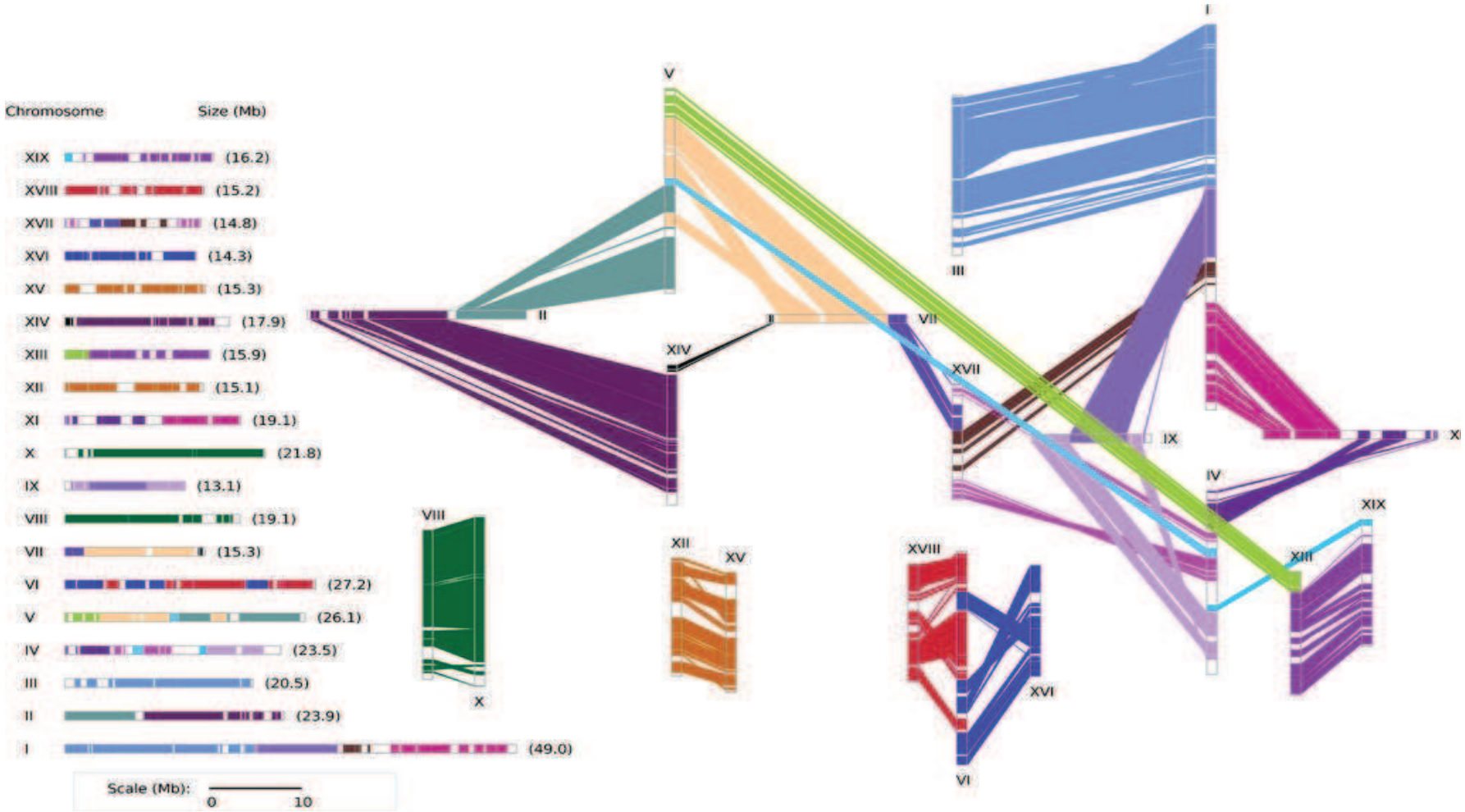


Figure I.9 : La version 2.2 de l'assemblage du génome du peuplier. Les 19 chromosomes sont désignés en chiffres romains, les couleurs identiques montrent les homologies de séquences entre chromosomes. À gauche, une représentation horizontale des chromosomes avec leur composition chimérique et leur taille en Mégabase (Mb). D'après (Tuskan et al. 2012).

génome. L'annotation de cette version est caractérisée par un nombre très important de transcrits 73013 et on retrouve 41 335 modèles de gènes (http://www.phytozome.net/poplar_er.php)

b) *Des cartes génétiques*

Depuis 1994, date de publication de la première carte génétique construite sur une famille interspécifique F2 *P. trichocarpa* x *P. deltoides* (Bradshaw et al. 1994), des cartes génétiques ont été développées sur 18 pédigrées de cartographie utilisant 13 espèces parentes différentes. Parmi ces 18 pédigrées, seuls 9 ont fait l'objet de phénotypage pour un ou plusieurs caractères d'intérêt à des fins de cartographie de QTL. Une présentation détaillée de ces pédigrées, cartes génétiques et QTL détectés est donnée dans le chapitre III.

c) *Des gènes candidats*

Les différentes catégories de gènes candidats couramment utilisées sont définies à partir des méthodes d'identification utilisées et reflètent des niveaux de connaissance et de validation différents de leur implication dans le contrôle de caractères d'intérêts. Ainsi, les approches de mutagenèse aléatoire suivies de génétique inverse permettent de définir des **candidats fonctionnels** pour lesquels seront mieux connus fonctions, produits et impacts directs sur le phénotype d'intérêt. Les approches de cartographie génétique, basées sur le polymorphisme et le déséquilibre de liaison donnent accès à des **candidats positionnels** présents dans l'intervalle de confiance des QTL détectés. La valeur de ces candidats dépend fortement de la taille de l'intervalle de confiance des QTL et de la qualité de l'annotation du génome disponible. Les études transcriptomiques basées sur l'étude des ARNm produits lors de la transcription du génome révèlent des **candidats expressionnels** pour des conditions définies (tissu, génotype, condition environnementale, temps donné d'une cinétique, etc.) à l'aide de puces à ADN, de PCR quantitative ou de séquençage d'ADN complémentaire.

Le statut d'espèce forestière modèle du peuplier permet aujourd'hui de disposer de nombreux gènes candidats appartenant à ces différentes catégories et relatifs à un grand nombre de caractères d'intérêt.

Des candidats fonctionnels :

Plusieurs outils de biologie moléculaire permettent d'accéder à la fonction du gène : la mutagénèse aléatoire (comme le Tilling et la mutagenèse insertionnelle), la mutagénèse dirigée et l'interférence par des ARN sens et antisens pour obtenir des mutants « gain of function » ou « loss of function ». La facilité d'obtenir des mutants chez le peuplier par rapport aux autres

arbres forestiers consolide sa position de modèle pour la compréhension de la fonction des gènes pour les espèces pérennes (Bradshaw et al. 2000; Taylor 2002; Jansson and Douglas 2007).

Au sein du genre *Populus,* différentes études ont porté sur le déterminisme moléculaire de la formation du bois et plus particulièrement sur la formation de la paroi secondaire et le fonctionnement du cambium (Sterky et al. 1998; Hertzberg et al. 2001; Plomion et al. 2001; Sterky et al. 2004; Vanholme et al. 2008 ; Vanholme et al. 2010). Les voies et les gènes impliqués dans la biosynthèse de la lignine, de la cellulose et de l'hémicellulose font l'objet d'un très grand nombre d'études et les rôles fonctionnels de plusieurs de ces gènes ont été validés par génétique inverse (extinction ou sur-expression) (Ralph et al. 2006 ; Leple et al. 2007) ou par l'analyse des mutants (Zhong et al. 2010). Des mutants transgéniques des gènes CAOMT (Caffeic Acid O-MethylTransferase), 4-coumarate : coenzyme A ligase et CCR (Cinnamoyl-CoA Reductase) impliqués dans la voie de biosynthèse des lignines ont été obtenus par interférence ARN sens et antisens (Tsai et al. 1998; Hu et al. 1999; Leple et al. 2007). Une réduction des transcrits de 3 à 4 % dans les mutants CCR provoque ainsi une réduction de 50% du taux de lignine total chez le peuplier transgénique (Leple et al. 2007). En 2000, Steve Strauss a suggéré une stratégie ambitieuse pour la découverte de gènes grâce à la génération de 50000 mutations dominantes (« gain of function ») avec l'insertion des T-DNA (Bradshaw et al. 2000). La mutagenèse insertionnelle qui permet aussi d'obtenir des mutations dominantes est utilisée aussi pour identifier des gènes responsables de fonctions données suite au screening de phénotypes engendrés. Une tail-PCR effectuée sur les fragments insérés, permet de révéler les gènes à proximité du site d'insertion. L'analyse de 1000 lignées transformées après génération de 1800 événements d'insertion a permis de monter que 2,4% des plantes générées ont un phénotype anormal au niveau des feuilles, des tiges et de la taille (Harrison et al. 2007). Cette étude a permis de révéler un nouveau membre de la famille des facteurs de transcription « Lateral Organ Boundaries Domain ». (Pta) *LBD1*est impliqué dans la régulation de la croissance secondaire du peuplier (Yordanov et al. 2010; Busov et al. 2011).

L'analyse de profils protéomiques fournit également des listes de gènes candidats fonctionnels. Ainsi des profils protéiques différents ont été obtenus pour le bois tendu et le bois opposé de deux génotypes de peuplier et des SAAPs (single amino acid polymorphism) ont été associées à ces différences (Abraham et al. 2012). Récemment, une approche haut débit appelée « shotgun proteome » a permis de déceler 6000 protéines impliquées dans la voie de biosynthèse des parois dont un grand nombre de protéines jusqu'à maintenant non identifiées (Kalluri et al. 2009).

Tableau I.4 : Liste des études expressionnelles de gènes candidats associés aux mécanismes de défense conduites chez le peuplier. D'après (Duplessis et al. 2009).

Famille	Nb. de gènes	Description	Références
MAPK (mitogen-activated protein Kinase)	21 MAPK	MAPKs sont activés par les chitosanes et par les éliciteurs des champignons *Septoria musiva*, *Melampsora medusae, Xanthomonas campestris* et *Xanthomonas populi*	(Hamel et al. 2005; Hamel et al. 2006; Nicole et al. 2006)
MAPKK (mitogen-activated protein Kinase Kinase)	16 MAPKK		
Biosynthèse de lignine	95	Après inoculation par *Entoleuca mammata* l'activité de la biosynthèse de la lignine augmente.	(Bucciarelli et al. 1998; Chen et al. 2000; Hamberger et al. 2007)
Métabolisme des flavonoïdes	48	Après une infection par *Melampsora medusae* le niveau des transcrits des gènes de la biosynthèse des flavonoïdes augmente, activant ainsi la synthèse des proanthocyanidines	(Tsai et al. 2006; Miranda et al. 2007)
Kunitz trypsin inhibitors	22	Après une infection par *Melampsora*, le niveau des transcrits des gènes du groupe C des KTI augmente	(Major and Constabel, 2008; Azaiez et al. 2009)
Peroxiredoxins	9	Le taux de protéines est altéré après une infection avec différentes races de *Melampsora larici-populina*	(Rouhier et al. 2004; Navrot et al. 2006; Gama et al. 2007)
Glutathione peroxidases	6		
13-lipoxygenase, type 2	3	Le niveau des transcrits de ces gènes augmente après : une infection par *Marssonina brunnea*, une blessure ou un traitement avec du jasmonate	(Cheng et al. 2006)
micro-ARN	234 appartenant à 48 familles	Sept cibles miARN de gènes de résistance ont été identifiées et 8 cibles sont validées expérimentalement	(Lu et al. 2005; Barakat et al. 2007; Lu et al. 2008)

De par leur rôle particulier identifié, les miARN (micro ARN) représentent des candidats fonctionnels potentiels. Ce sont de petits ARN non codants d'environ 21 nucléotides chez les Angiospermes. Chez les plantes, les gènes qui leur sont associés sont généralement présents dans les régions inter-géniques et possèdent leur propre unité de transcription. La transcription d'un gène codant un miARN conduit à la formation d'un miARN dit primaire (pri-miARN) qui adopte une conformation particulière de type tige boucle. Ce précurseur subira une maturation dans le noyau avant son export vers le cytoplasme où il va se lier avec la protéine argonaute AGO1 pour former le complexe RISC (RNA Induced Silencing Complex). Ce complexe RISC, guidé par le miARN monocaténaire mature, induit alors un silencing post-transcriptionnel des gènes cibles par dégradation des ARNm ou répression de leur traduction (Jones-Rhoades et al. 2006; Zhang et al. 2007). La présence de nombreux miARN chez le peuplier qui n'ont pas d'analogues chez *Arabidopsis thaliana* (Barakat et al. 2007) suggère une probable implication des miARN dans les processus spécifiques à la biologie des arbres. Par exemple, les miARN interviennent dans la formation du bois de tension (Lu et al. 2005). Les miARN interviennent de façon plus générale dans les processus de développement de la plante et de sa croissance ou en réponse à différents stress ou interviennent dans les mécanismes de défense chez le peuplier (Lu et al. 2005; Lu et al. 2008).

Des candidats expressionnels

Le séquençage du transcriptome du peuplier (EST) a permis de révéler un nombre important de gènes candidats expressionnels.

La première bibliothèque a été créée en Suède avec pour objectif de mieux comprendre les mécanismes moléculaires impliqués dans la formation des parois cellulaires. L'étude des EST exprimés dans un tissu particulier tel que le bois de tension, tissu est riche en cellulose, a permis de révéler un taux important de transcrits de sucrose synthase et d'autres candidats d'intérêt (Dejardin et al. 2004 ; Andersson-Gunneras et al. 2006 ; Zhou et al. 2009; Zhong and Ye 2010; McCarthy et al. 2010; Tsai et al. 2011).

Plusieurs études du transcriptome s'intéressent aux réponses de différentes espèces de peuplier à des stress abiotiques (déshydratation, froid, salinité, chaleur, exposition à l'ABA et H_2O_2). Ainsi sur la seule espèce *P. nigra,* 4500 gènes spécifiquement liées à ces réponses ont été identifiés (Nanjo et al. 2004). Sur *P. euphratica*, 7841 unigènes ont été obtenus suite au séquençage de 17 bibliothèques d'EST associés à des réponses à la sècheresse et au stress salin (Brosche et al. 2005). Grâce aux nouvelles technologies de séquençage (454 DNA sequencing technology), l'Oak Ridge National Laboratory (ORNL) et le DOE joint Genome Institute ont séquencé 6

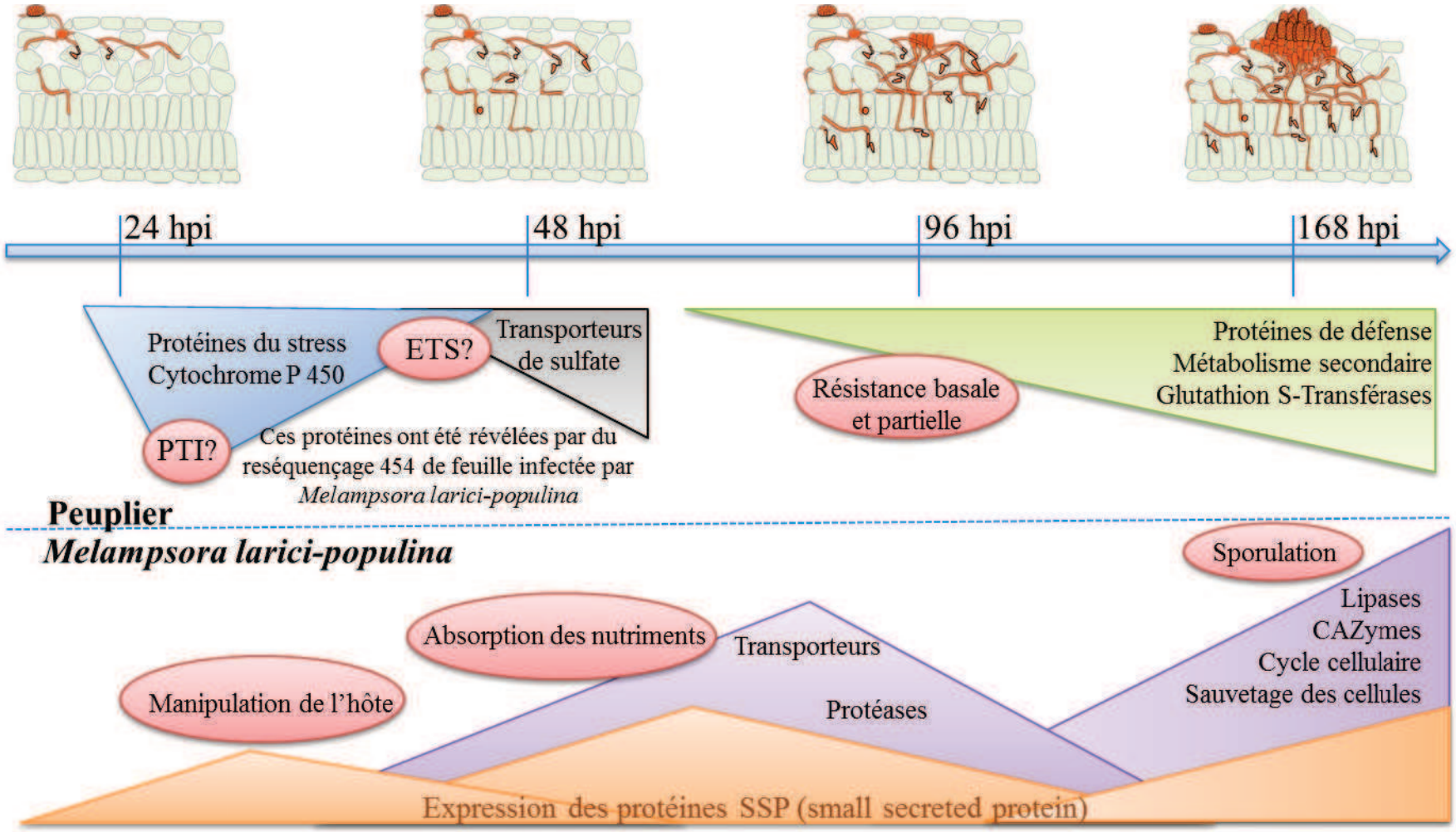


Figure I.10 : Représentation schématique des principales régulations du transcriptome du peuplier et de *Melampsora larici-populina* dans le cadre de l'interaction compatible. hpi : heure après inoculation, PTI: PAMP-triggered immunity; ETS: effector-triggered susceptibility. D'après (Hacquard et al. 2011).

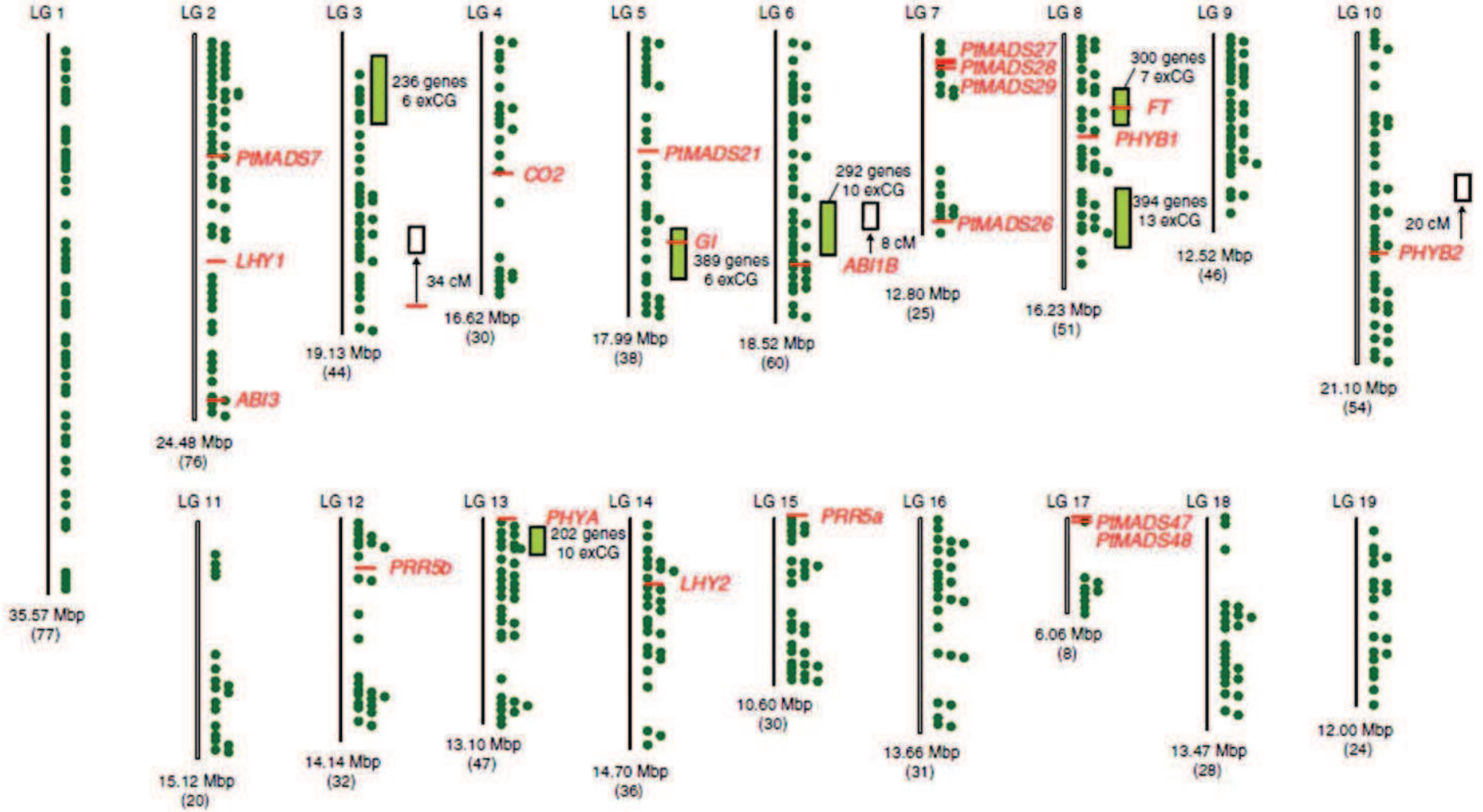


Figure I.11 : Représentation schématique sur les 19 chromosomes du peuplier des QTL (boites vertes et blanches), des gènes candidats fonctionnels (en rouge) et des gènes candidats expressionnels (points verts) associés à l'arrêt de croissance chez le peuplier. D'après (Rohde et al. 2011)

bibliothèques d'ADNc provenant de feuilles de *P. deltoides* récoltées en conditions de stress hydrique. L'analyse de l'ontologie des gènes a révélé un enrichissement de gènes de résistance à la sécheresse (Yang et al. 2009).

Concernant des caractères de résistances à des contraintes biotiques, des études d'expression ponctuelles sur quelques ARNm ont d'abord permis de révéler des protéines 14-3-3 pour la transmission du signal ou encore une activité de la PAL (Phénylalanine ammonia-lyase) accrue lors d'une infection (Lapointe et al. 2001). Des études transcriptomiques au niveau du génome entier ont permis de révéler des candidats expressionnels fortement exprimés lors de l'infection de feuilles de peuplier par des champignons pathogènes du genre *Melampsora* (Miranda et al. 2007; Rinaldi et al. 2007; Azaiez et al. 2009; Duplessis et al. 2011), dans la réponse à des attaques d'insectes défoliateurs (Ralph et al. 2008 ; Philippe et al. 2009) ou dans la réponse à l'herbivorie (Babst et al. 2009; Philippe et al. 2010). Deux articles de synthèse récents reportent des listes de gènes impliqués dans les mécanismes de défense du peuplier à des bioagresseurs (**Tableau I.4**), une liste de facteurs de transcription appartenant à différentes familles (WRKY, bZYP, ERF, MYB, JAZ) (Duplessis et al. 2011) et une liste de candidats plus spécifiquement impliqués dans l'interaction compatible Peuplier- *Melampsora larici-populina*, agent de la rouille foliaire (**Figure I.10**).

Des candidats positionnels :

Plusieurs études de cartographie de QTL ont été menées chez le peuplier pour l'identification des régions du génome responsable de la variation de caractères d'intérêt tels que la croissance, la phénologie, la qualité du bois, l'efficience d'utilisation de l'eau ou la résistance à différents agresseurs (cf chapitre III). L'existence d'une séquence de référence annotée chez le peuplier permet de définir des listes de gènes candidats positionnels présents au sein des intervalles de confiance des QTLs. Compte tenu de la taille de ces intervalles de confiance, ces listes comportent souvent une centaine voire un millier de gènes ! (Fabbrini et al. 2012).

Deux stratégies sont utilisées pour réduire de façon significative ces listes de candidats :

- la première vise à croiser pour une même région génomique, liste de candidats expressionnels et fonctionnels connus. Elle a été notamment mise en œuvre pour préciser une liste de gènes candidats impliqués dans le contrôle de l'arrêt de croissance et la formation du bourgeon terminal (**Figure I.11**)
- la deuxième utilise des outils bioinformatiques de recherche d'enrichissement significatifs de modèles de gènes particuliers dans ces régions d'intérêt. Elle a été mise en œuvre pour préciser

Tableau I.5 : Ressources génomiques du peuplier modifié et actualisé d'après (Brunner et al. 2004)

Institution	URL	Contenu	Références
International Populus Genome Consortium, Oak Ridge National Laboratory, USA	http://www.ornl.gov/ipgc/	Coordination de la recherche internationale, base de données SSR , protocoles, liens	
Swedish Populus Genome Project	http://poppel.fysbot.umu.se/	Base de données EST, microarrays	
PoplarDB/Lignome, INRA	http://mycor.nancy.inra.fr/poplardb/	Base de données EST	
Oak Ridge National Laboratory Computational Biology, USA	http://genome.ornl.gov/poplar_chloroplast/index.html	Annotated Populus chloroplast genome sequence	
Genome Canada, University of British Columbia and Laval University	http://www.bcgsc.ca/gc/poplar; http://www.arborea.ulaval.ca/en/	Cartes physiques, séquences BAC-end, ESTs, microarrays, SSR, transgéniques et mutants	
Dendrome, US Forest Service and University of California, Davis	http://dendrome.ucdavis.edu/	Bases de données et liens pour les arbres forestiers, annuaire des scientifiques.	
Michigan Technological University Plant Biotechnology Research Center	http://forest.mtu.edu/research/pbrc/	Base de données EST, cellulose and lignine.	
Oregon State University Tree Genetic Engineering Research Cooperative	http://people.forestry.oregonstate.edu/steve-strauss/tbgrc-tree-biosafety-and-genomics-research-cooperative	Service de transformation génétique, collection de mutants.	
Department of Energy's Joint Genome Institute and the Center for Integrative Genomics	http://www.phytozome.net	Base de données des génomes de plantes. C'est la base de donnée référence pour le peuplier.	(Goodstein et al. 2012)
Sweden defence research Agency and University of Umea (Umea Plant Science Center)	http://www.popgenie.org/	Intégration de données du génome, transcriptome, synthénie et QTL.	(Sjodin et al. 2009)
Center for Bioinformatics of Peking University and The Key Laboratory of Forest Genetics and Gene Engineering, Nanjing Forest University	http://dptf.cbi.pku.edu.cn/	DPTF : Database of Poplar Transcription Factor, 2576 séquences de facteurs de transcription	(Zhu et al. 2007)
Forestry and Forest Products research Institute and Riken plant science center (Japan)	RPOPDB : http://rpop.psc.riken.jp/index.pl	Informations sur les séquences d'ARNm de 10 espèces *Populus*	
Computational & Systems Biology Laboratory University of Georgia	RepPop : http://csbl.bmb.uga.edu/~ffzhou/RepPop/	Tous les éléments répétés dans génome du peuplier.	(Zhou and Xu 2009)
Department of genetics at the university of Georgia	AspenDB: http://aspendb.uga.edu/index.php	Annotation des puces peuplier (Affymetrix, Agilent v1 & v2)	(Tsai et al. 2011)
GabiPD, Federal Ministry of Education and Research and Max-Planck-Institute of molecular Plant	http://www.gabipd.org/	9854 séquences nucléotidiques de *P. tremula* et *P. tremula* x *P. tremuloides*	(Riano-Pachon et al. 2009)
Genbank (NCBI) and DDBJ EMBL EBI (European Bioinformatics Institute)	http://www.ncbi.nlm.nih.gov/ http://www.ddbj.nig.ac.jp/ http://www.ebi.ac.uk/	3822 articles, 164595 séquences nucléotidiques, 455249 EST, 575 298 protéines et 45024 modèles de gènes (Genbank) ; 156336735 bp et 109 455 entrées (DDBJ, position 89) ; 629 330 séquences nucléotidiques, 165 227 protéines, 41 378 modèles de gènes (EnsemblPlants)	Genbank, DDBJ, EMBL ont fusionné, ils ont pratiquement les mêmes ressources
InterPro (Prosite, Pfam, Prints, Prodom, SMART, TIGRFAMs, HAMAP, Panther, PIRSF, Superfamily, Gene3D)	ttp://www.ebi.ac.uk/interpro/	Toutes les ressources EMBL peuplier	New Interpro beta (28 aout 2012)
Uniprot	http://www.uniprot.org/uniprot/?query=organism:3694+keyword:1185	41681 références protéome peuplier	

Tableau I.5 (suite)

Institution	URL	Contenu	Références
The Computational biology and Functional Genomics Laboratory	http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=poplar	ESTs, séquences génomiques, annotations fonctionnelles, voies métaboliques, gene ontology, prédiction d'oligomères.	
GRAMENE (USDA-ARS at cold spring Harbor laboratory and cornell university, Oregon state university	http://www.gramene.org/Populus_trichocarpa/Info/Index	alignement des sondes du Poplar Genome Array Affymetrix GeneChip sur le génome, séquences nucléotidiques, plant metabolic pathway(poplarCyc), Blast et biomart	
BIO5 Institute, University of Arizona	chromDB : www.chromdb.org	654 gènes du peuplier avec la séquence nucléotidique, transcrits, ORF et protéine.	(Gendler et al. 2008)
Iowa State University, USDA-ARS Corn Insects and Crop Genetics, University of South Dakota	PlantGDB : http://www.plantgdb.org/PtGDB/	plus de 27 espèces végétales, données nucléotidiques, comparaison de génomes	(Duvick et al. 2008)
University of Manchester and The Welcome Trust Sanger Institute	miRBase : http://www.mirbase.org/	328 séquences de miRNA spécifiques au peuplier (localisation et annotation).	(Griffiths-Jones et al. 2008)
Laboratory of Plant Physiology and Biochemistry and laboratory for Agricultural Biotechnology, College of Biological Sciences, China Agricultural University	PMRD: plant microRNA database http://bioinformatics.cau.edu.cn/PMRD/	Plus de 2780 séquences de miRNA prédis chez *P. trichocarpa* et d'autres espèces du peuplier.	(Zhang et al. 2010)
INRA URGV, UAGPF, GAFL	http://urgv.evry.inra.fr/FLAGdb	Flag dB++ est une base de données regroupant 5 espèces végétales dont le peuplier.	(Derozier et al. 2011) avant l'intégration de la tomate
URGI - Unité de Recherche Génomique Info (a research unit in genomics and bioinformatics, INRA).	http://urgi.versailles.inra.fr/Species/Populus/Resources	13 espèces végétales dont le peuplier (séquences nucléotidiques, annotations, SSR, SNPs, cartes)	

une liste de gènes candidats impliqués dans le contrôle de la croissance et de l'efficience d'utilisation de l'eau (Monclus et al. 2012).

d) *Des bases de données (ADN, ARN, microARN, protéines)*

Le statut de modèle du peuplier a permis le développement de plusieurs bases de données spécifiques au peuplier ou des bases de données générales intégrant les données du peuplier. Brunner et al (2004) ont recensé plusieurs sites web dédiés au peuplier, dont une liste remise à jour pour 2012 est présentée dans le (**Tableau I.5**)

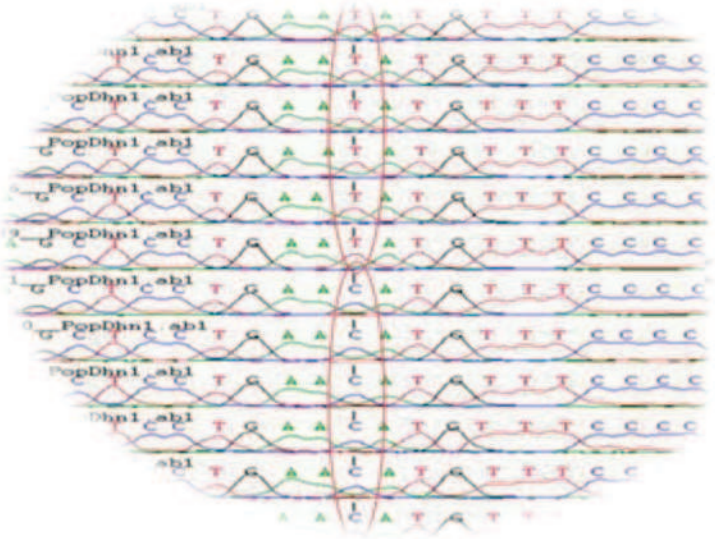
Une source importante de données se trouve sur le site de l'ORNL qui a coordonné les travaux du consortium international du génome du peuplier. La base de données Phytozome (http://www.phytozome.net/Phytozome_resources.php) est la base de référence pour les ressources génomiques végétales. Elle regroupe les données de 31 espèces végétales différentes dont le peuplier. On retrouve les informations sur les séquences génomiques, les transcrits, les annotations de gènes, les protéines et les arbres phylogénétiques.

Cette base de données s'appuie sur plusieurs sites web et bases de données internationales accessibles depuis ce lien. Parmi ces bases de données, on retrouve celles développées plus particulièrement sur des espèces forestières et le peuplier (dendrome, TreeGenes EST, PopulusDB expression et PopGenie). Phytozome a soumis ces données aux grandes bases de données internationales. On retrouve ainsi des ressources génomiques peuplier sur la DDBJ/EMBL/GenBank (DNA Data Bank du Japan (DDBJ), EMBL-Bank en Europe et la GenBank (NCBI) aux États Unis respectivement). Plusieurs bases de données ne se limitant pas au peuplier permettent la comparaison entre génomes et des études plus profondes et larges sur la génomique des végétaux (PlantGDB, GRAMENE, GabiPD, etc.). Flagdb++ (France) qui a intégré les données relatives au peuplier permet la compréhension du rôle biologique des gènes dans un contexte large: une famille multigénique, un environnement topologique, et/ou un réseau fonctionnel. Cette base intègre les annotations structurales et fonctionnelles internationales, les EST de différentes espèces végétales, des prédictions de nouveaux gènes, des mutants tags , des familles de gènes, des protéines, des motifs de données du transcriptome, des séquences répétées, des sondes pour les approches génomiques (puces à ADN, études de synthénie, criblage de banque BAC, RT-PCR, la découverte de SNP, ...), des structures secondaires, des modèles 3D et des phénotypes mutants. Deux bases de données spécifiques aux miARN ont été développées : miRbase est générale alors que PMRD est une base de données dédiées aux plantes.

Les données de séquençage sont disponibles publiquement dans plusieurs sites alors que les données de génotypage sont rarement accessibles. Les SNP développés dans ce travail de thèse seront intégrés dans la base de données GnpSnp de la plateforme URGI de l'INRA.

Chapitre II

- Développement d'une boite à outils moléculaires pour la cartographie génétique et les études d'association chez *P. nigra*



Chapitre II : Développement d'une boite à outil moléculaire pour la cartographie génétique chez *P. nigra*

A. État de l'art

En profitant des ressources génomiques disponibles chez le peuplier, cette thèse vise à développer une boîte à outils moléculaires permettant de balayer l'ensemble du génome du peuplier noir (*P. nigra* L.) afin d'identifier les régions génomiques, les gènes candidats et les polymorphismes impliqués dans des caractères d'intérêt pour la production de biomasse en TCR et TtCR. Ainsi dans cette étude, nous souhaitons principalement développer des marqueurs à des fins de cartographie génétique. Ces marqueurs doivent donc être abondants, polymorphes, bien répartis sur tout le génome et doivent permettre un ancrage sur la séquence de référence. En outre, ce travail présente comme second objectif d'étudier le niveau et l'organisation de la diversité nucléotidique au sein de gènes candidats pour les caractères d'intérêt afin d'une part d'apporter des connaissances sur l'étendue du déséquilibre de liaison chez *P. nigra* et d'autre part de trouver d'éventuelles traces de sélection et ainsi affiner la liste de gènes candidats pour de futures études d'association. Le séquençage de gènes candidats dans un échantillon représentatif de la diversité génétique de l'espèce permet de répondre à ces objectifs.

1. Identification de polymorphismes SNP par reséquençage de gènes

Les SNP constituent la forme la plus abondante de variations héréditaires à l'échelle moléculaire (Gibson and Muse 2004). La mise en évidence de ces marqueurs a été faite sur l'humain lors de l'étude de sujets sains et malades portant des allèles différents d'un gène donné. Les efforts de séquençage du génome humain ont permis d'identifier plusieurs millions de SNP et de les génotyper (Frazer et al. 2007). Les marqueurs SNP présentent un polymorphisme souvent bi-allélique et sont distribués uniformément dans tout le génome. L'importance grandissante donnée aux SNP tient à leurs diverses applications (Ganal et al. 2009) : la construction de cartes génétiques à haute résolution, la cartographie de gènes candidats, les études d'associations, la génétique des populations (génomique des populations) et la phylogénétique. La séquence du génome entier de *P. trichocarpa* a permis de révéler un niveau d'hétérozygotie élevé chez le peuplier (1 SNP / 385pb), ce qui correspondrait à plus de 1,2 millions de SNP découvert sur un seul individu pour une taille de génome de 485 Mb (Tuskan et al. 2006; Drost et al. 2011).

La détection du polymorphisme SNP peut se faire par séquençage ou par des méthodes indirectes. Malgré les progrès récents de la génomique qui laissent entrevoir des possibilités de

reséquençage de génomes entiers, la plupart des travaux publiés et en cours chez les arbres forestiers concernent l'identification de SNP par séquençage de gènes candidats responsables de la variation des caractères d'intérêt (Neale and Ingvarsson 2008; Neale and Kremer 2011). Quant à elles, les techniques indirectes d'identification de SNP reposent sur l'observation de l'ADN après sa dénaturation. La présence d'un polymorphisme au sein du fragment analysé, dans un pool de séquences appartenant à différents individus, provoque la formation d'hétéroduplex d'ADN (molécules double brin formées par réappariement des deux brins issus d'allèles différents) lors de sa renaturation. Cette propriété a été la base du développement de la méthode d' « Ecotilling » (Comai et al. 2004). Elle consiste à mélanger l'ADN de l'échantillon à analyser avec un ADN de référence en quantité égale avec des amorces spécifiques portant des fluorescences différentes. Après amplification, les produits PCR sont dénaturés. Les homoduplex ou les hétéroduplex se forment alors après renaturation. Une enzyme spéciale repère les hétéroduplex et les coupe à l'endroit du polymorphisme. Les profils des fragments générés sont révélés par migration sur gel. Cette technique a été utilisée pour évaluer la diversité nucléotidique dans 9 gènes candidats pour la phénologie chez *P. trichocarpa* (Gilchrist et al. 2006) et dans 35 gènes candidats pour les propriétés du bois chez *P. tremuloides* (Kelleher et al. 2012). Si ces méthodes présentent l'avantage d'être moins coûteuses par rapport au séquençage, elles présentent l'inconvénient de ne pas fournir autant d'information.

Quelle que soit la technique d'identification choisie, pour maximiser la détection de SNP, il faut explorer au mieux la diversité génétique présente dans l'espèce d'étude. Pour cela, il est nécessaire de connaitre la structuration spatiale de cette diversité au sein des populations naturelles et à l'échelle de l'aire de distribution. Des études réalisées à l'aide de marqueurs neutres (6 à 7 SSR) chez le peuplier noir, ont montré une diversité élevée au sein des populations naturelles qui sont peu différenciées à l'échelle d'un bassin versant (Lefèvre et al. 2002 ; Smulders et al. 2008b). À l'échelle de la partie Ouest de l'aire de distribution de l'espèce, ces études ont par ailleurs montré une diversité génétique particulièrement élevée dans les populations françaises avec une légère différenciation entre les bassins versants (Lefèvre et al. 2002 ; Imbert and Lefevre 2003; Storme et al. 2004; Smulders et al. 2008b). En revanche, une différenciation plus importante a été observée sur des caractères adaptatifs suggérant des adaptations locales (Chamaillard et al. 2011; Rohde et al. 2011). Dans ce contexte, le choix d'un panel de découverte ne nécessitera pas l'échantillonnage d'un très grand nombre de populations couvrant l'ensemble de l'aire de distribution, mais devra plutôt privilégier l'échantillonnage d'individus non apparentés d'origine française variée.

Tableau II.1 : Patrons de diversité nucléotidique chez les salicacées. $\mathbf{N_{gènes}}$: nombre de gènes analysés, $\mathbf{N_{pop}}$: nombre de populations analysées, $\mathbf{N_{gam}}$: nombre de gamètes (généralement, nombre d'individus x 2), $\mathbf{S_{kb}}$: nombre de SNPs dans 1000 pb de séquence, $\boldsymbol{\pi}$: diversité nucléotidique, $\boldsymbol{\pi_S}$: diversité synonyme, $\boldsymbol{\pi_N}$: diversité non synonyme, $\mathbf{LD_{0,2}}$: distance entre paires de SNP correspondant à r^2=0,2, $\mathbf{F_{ST}}$: la différentiation entre populations.

Espèces	$N_{gènes}$	N_{pop}	N_{gam}	S_{kb}	π	π_S	π_N	$LD_{0,2}$	F_{ST}	Références
P. nigra	9	-	48	38,7	0,007	0,0107	0,0046	300	-	(Chu et al. 2009)
	31	-	24	8,2	0,0024	0,0029	0,0009	<300	-	G. Zaina comm perso
	6	-	24	10,3	0,0008	-	-	-	-	(Fladung and Buschbom 2009)
	1	5	720	20,1	0,0012	-	-	<50pb	-	(Marroni et al. 2011b)
P. trichocarpa	9	-	78	7,7	0,0018	0,0029	-	-	-	(Gilchrist et al. 2006)
	10	3	94	3,69	0,0014	-	-	3 à 6 kb	0,013 - 0,203	(Slavov et al. 2012)[a]
	9	5	60	6,02	0,003	-	-	<400	-	(Ismail et al. 2012)
P. balsamifera	3	8	204 - 210	10,67	0,0025	0,0033[b]	0,0034	-	0,256[c]	(Breen et al. 2009)
	11	7	188	18,7	0,002	0,003	0,0008	-	0,053	(Keller et al. 2010)
	460	15	30	10,5	0,0026	0,0045	0,0012	-	-	(Olson et al. 2010)
	27	-	16 - 48	8,48	0,0019	0,003	0,0007	-	-	(Keller et al. 2011)
	9	5	60	17,1	0,005	-	-	250	-	(Ismail et al. 2012)
P. tremula	5	4	48	-	0,0111	0,022	0,0059	<100	0,117	(Ingvarsson 2005b)
	6	4	44	68	0,0144	0,0207	0,0117	-	0,107	(Ingvarsson 2005a)
	1	4	48	39,3	0,0061	0,0085	0,003	-	0,045	(Ingvarsson et al. 2006)
	76	-	24 - 38	19	0,0042	0,012	0,0017	<100	-	(Ingvarsson et al. 2008)
	6	-	8	9	0,0022	-	-	-	-	(Fladung and Buschbom 2009)
	8	-	22 - 54	26,42	0,0077	0,0147	0,0041	-	-	(Bernhardsson and Ingvarsson 2011)
	7	12	232	-	-	-	-	-	0,021 et 0,064	(Bernhardsson and Ingvarsson 2012)
	9	3	60	18,3	0,007			<400	-	(Ismail et al. 2012)
P. tremuloides	6	-	24	12,1	0,0011	-	-	-	-	(Fladung and Buschbom 2009)
	35	4	128	18	0,0006 - 0,0126	-	-	20	0,004	(Kelleher et al. 2012)
P. alba	6	-	16	3,4	0,0004	-	-	-	-	(Fladung and Buschbom 2009)
P. deltoides	6	-	16	4,7	0,0006	-	-	-	-	(Fladung and Buschbom 2009)
S. viminalis	18	-	48	24,2	0,0049	0,0067	0,0014	4000	-	(Berlin et al. 2011)
S. schwerinii	18	-	48	27,4	0,007	0,0095	0,003	1000	-	(Berlin et al. 2011)

[a] dans cette étude, le calcul de π est basé sur du reséquençage de 10 gènes candidats chez 47 arbres, le calcul du DL sur16 arbres génomes entiers et dans 3704 gènes candidats (120 arbres), et le calcul du Fst sur 22 280 loci (SNP) dans 10 populations.

[b] familles de demi-frères.
[c] calculé à partir des données de re-séquençage de 15 individus *P. trichocarpa* non apparentés.

2. Niveau et organisation de la diversité nucléotidique chez les peupliers

La diversité nucléotidique (π) a été définie pour mesurer le niveau de polymorphisme au sein d'une population. Elle estime le nombre moyen de différences nucléotidiques par site entre toutes paires de séquences choisies au hasard dans un échantillon de la population (Nei and Li 1979). À l'échelle d'un individu, la diversité nucléotidique est comparable à l'hétérozygotie attendue (H_e). Il est possible de décomposer la diversité nucléotidique totale en diversité non synonyme (π_N) et synonyme (π_S), selon que le polymorphisme conduise ou pas à un changement d'acide aminé dans le produit du gène. Cette distinction permet notamment d'inférer l'amplitude et le signe d'une sélection naturelle agissant sur les gènes étudiés.

Le niveau de diversité nucléotidique a été étudié chez 7 espèces de peupliers dans 15 études différentes sur 1 à 460 gènes ou fragments de gènes (**Tableau II.1**). Des niveaux variables de diversité ont été rapportés aussi bien en intra qu'en interspécifique. En effet les valeurs de π varient entre 0,0004 chez *P. alba* et 0,0144 pour *P. tremula*. Plusieurs faibles valeurs s'expliquent par la petite taille du panel de découverte utilisé. Ces valeurs restent néanmoins comparables aux niveaux de diversité nucléotidique rapportés pour 2 espèces de saules (**Tableau II.1**). Ces études rapportent un nombre de polymorphisme élevé avec entre 3,4 et 68 SNP pour 1000 pb ce qui correspond environ à 1 SNP tous les 15 à 300 pb. Parmi ces études, celles qui distinguent la diversité synonyme et non synonyme ont toutes rapporté, comme attendu, des valeurs de diversité synonyme supérieures ou égales aux valeurs de diversité non synonyme.

Le déséquilibre de liaison (DL) est défini comme l'association non aléatoire entre allèles à des locus différents. Il traduit une déviation de la fréquence de certaines combinaisons d'allèles ou de marqueurs génétiques par rapport à l'attendu aléatoire basé sur leurs fréquences (appelée D). Il peut être mesuré de différentes manières, la plus couramment utilisée est le r^2 défini de la façon suivante : $r^2 = D/\sqrt{f_A f_B (1-f_A)(1-f_B)}$, avec $D = f_{AB} - f_A f_B$, f_A et f_B les fréquences des 2 allèles A et B. Le niveau de DL est caractéristique d'une population donnée. Il est influencé par différent facteurs : la liaison physique, le taux de recombinaison, les pressions évolutives (mutation, sélection, dérive et migration).

Le DL est un paramètre très important dans les études de cartographie QTL et de génétique d'association (Rafalski 2002; Flint-Garcia et al. 2003; Slatkin 2008). C'est le principe de base sur lequel repose la cartographie QTL : on ne détecte des QTLs que lorsqu'il y a un DL entre les marqueurs de la carte génétique et le QTL. Ce DL est à son maximum après une génération (1 événement de méiose) et devient très faible au fur et à mesure des générations

Tableau II.2 : Nombre de gènes communs entre les études de diversité dans le genre *Populus* citées dans le tableau II.1.

	P. balsamifera	***P. trichocarpa***	***P. nigra***	***P. deltoides***	***P. tremula***	***P. tremuloides***	***P. alba***	***Gènes***
1 espèce	516	5	7	0	26	28	0	sans *a priori*
2 espèces	2	1	1	0	1	1	0	Adh 1, POPTR_0010s07330, Poptr4CL3
3 espèces	7	7	0	0	7	0	0	PTCBF2, TI-4, Dhn, GI, WIN3, GAPDH, TI-3
4 espèces	2	2	0	0	2	2	0	PAL1, phyB2
5 espèces	1	1	4	3	4	4	3	CBF1, C3H1, PPO, GA20ox
6 espèces	0	2	2	2	2	2	2	TB1, CAD-like
7 espèces	1	1	1	1	1	1	1	LFY
Nb total de gènes	529	19	15	6	43	38	6	

Chez les peupliers, le DL décroit à très courte distance physique : le r^2 atteint 0,2 ($LD_{0,2}$) au bout de 20 pb chez *P. tremuloides* dans un panel de 64 individus et 35 gènes candidats (**Tableau II.1**). Ce DL est variable au sein des autres espèces du genre *Populus* mais il reste faible dans toutes les études réalisées que ce soit à partir des SNP issus du séquençage de gènes ou encore d'ecotilling ($LD_{0,2}$ < 400 pb). Toutefois, ce DL ne décroit à 0,2 qu'à partir de 3 kb chez *P. trichocarpa* dans un panel de 16 génomes entier et dans 3704 gènes candidats appartenant à ces génomes (Slavov et al. 2012). Ce DL relativement élevé par rapport aux autres références est toutefois comparable aux estimations obtenues chez 2 espèces de saule où le DL se maintient à plus de 0,2 jusqu'à 1 kb chez *S. schwerinii* et même 4 kb chez *S. viminalis* (Berlin et al. 2011). Chez les conifères, où le DL décroit généralement très rapidement, il peut se maintenir à plus de 0,2 jusqu'à 1,5 kb comme rapporté chez le pin *taeda* (Neale and Savolainen 2004). Les raisons pouvant expliquer les divergences observées entre espèces de peuplier restent obscures. Il est très difficile de faire la part de l'influence du taux variable des recombinaisons le long du génome, de celles de la structuration génétique et de l'histoire évolutive des populations et espèces échantillonnées (Ingvarsson 2008; Olson et al. 2010; DiFazio et al. 2011).

Le faible DL mis en évidence dans les populations naturelles d'arbres forestiers pourrait rendre les études d'association très précises à condition d'avoir une densité de marquage énorme sur tout le long du génome, ce qui est pour l'instant difficile malgré l'émergence phénoménale des nouvelles techniques de séquençage et de génotypage. Pour contourner ce problème, les associations sont faites sur des SNP détectés dans des gènes candidats validés par des études préalables en génétique inverse ou en transcriptomique (Pflieger et al. 2001), ce qu'on définit comme des gènes candidats.

Les différentes études sur la diversité moléculaire chez les peupliers se sont soit focalisées sur quelques gènes candidats impliqués dans la résistance aux rouilles et aux herbivores, la tolérance à la sécheresse et au froid, la croissance, la biosynthèse des hormones, la qualité du bois et la phénologie (Ingvarsson 2005a ; 2005b ; Gilchrist et al. 2006 ; Breen et al. 2009; Chu et al. 2009; Fladung and Buschbom 2009 ; Keller et al. 2011; Bernhardsson and Ingvarsson 2011; Ismail et al. 2012; Kelleher et al. 2012), soit sur de nombreux locus sans *a priori* sur les fonctions des gènes (Ingvarsson 2008; Keller et al. 2010; Olson et al. 2010). Il y a relativement peu de gènes candidats en commun entre toutes ces études (**Tableau II.2**).

Sur les bases acquises de la littérature (avant 2010) nous avons rassemblé une liste de gènes candidats positionnels, expressionels et fonctionnels pour les caractères d'intérêt pour la culture de peuplier en TCR et TtCR. Pour maximiser nos chances de découvrir des SNP, nous avons reséquencé 665 fragments de gènes, sur un panel de 21 individus *P. nigra* non apparentés

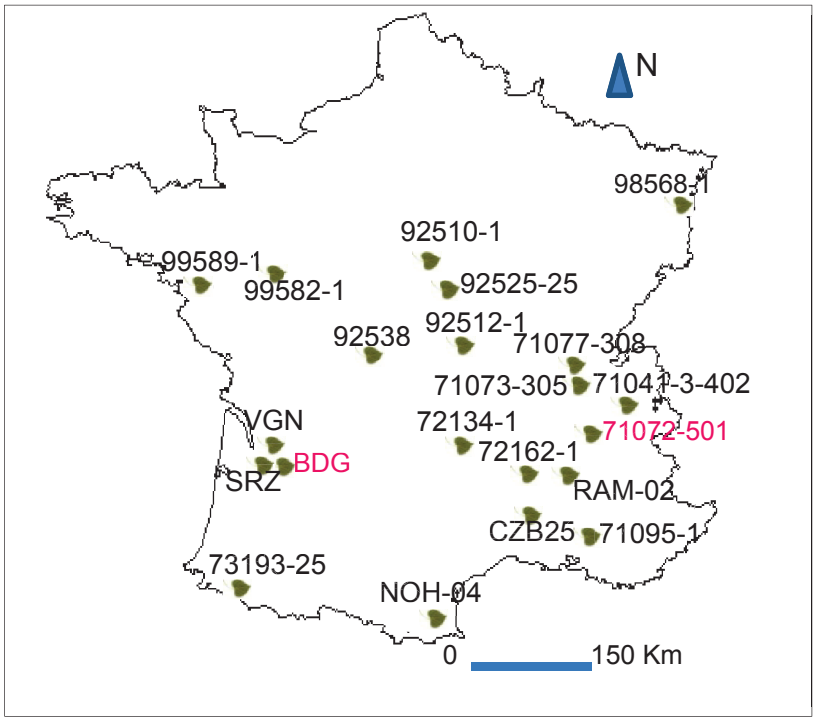


Figure II.1 : Origines géographiques des 21 individus *Populus nigra* utilisés pour le re-séquençage de gènes. En rouge, les deux parents de la population de cartographie (la femelle 71072-501 et le mâle BDG).

échantillonnés dans des populations naturelles françaises présentes dans différents bassins versants. Nous avons évalué le niveau et l'organisation de la diversité nucléotidique dans ces fragments de gènes. Il s'agit d'une première étude conduisant à l'estimation de la diversité nucléotidique dans un nombre important de gènes candidats chez le peuplier noir.

B. Matériel et méthodes

1. Matériel végétal

En 2008, 21 génotypes *P. nigra* ont été sélectionnés au sein de populations naturelles françaises présentes dans 6 bassins versants (Loire, Garonne, Adour, Têt, Rhône et Rhin) pour constituer un panel de découverte (**Figure II.1**). Huit individus sont impliqués comme parents dans un plan de croisement factoriel intraspécifique 5x4. Deux de ces 8 individus (la femelle 71072-501 et le mâle BDG) sont les parents d'une famille de plein-frères de 324 génotypes clonés utilisée à des fins de cartographie génétique et de détection de QTL dans le présent travail de thèse (chapitres III, IV et V). Des échantillons de feuilles juvéniles ont été récoltés en 2008 dans le parc à pieds mère installée à la pépinière forestière de l'État de Gueméné-Penfao pour effectuer des extractions d'ADN.

2. Choix des gènes candidats

Dans une première étape, une liste de 9241 gènes a été établie et rassemble (**Figures II.2 et II.3**) :

- 5428 gènes candidats positionnels liés à l'efficience de l'utilisation de l'eau (WUE) issus de travaux de cartographie QTL (Monclus et al. 2012)
- 413 gènes candidats pour la formation et les propriétés du bois issus à la fois d'études expressionnelles et fonctionnelles. Il s'agit pour les candidats expressionels de gènes exprimés dans la formation du cambium de peuplier (Goué 2005). Les gènes candidats fonctionnels sont quant à eux liés à la synthèse des microtubules, au métabolisme primaire de sucres, de la cellulose et de la lignine qui ont été mis en évidence lors de la formation des parois secondaires lignifiées (Dejardin et al. 2004)
- 60 gènes candidats expressionnels et positionnels impliqués dans la résistance la rouille foliaire à *Mlp*. Les candidats expressionnels étaient différentiellement exprimés dans les premières étapes d'une réaction incompatible du cultivar « Beaupré » (Rinaldi et al. 2007). Les candidats positionnels correspondent au locus impliqué dans le contrôle de la taille des sores chez le peuplier hybride et hérité de *P. trichocarpa* (Bresson et al. 2011)

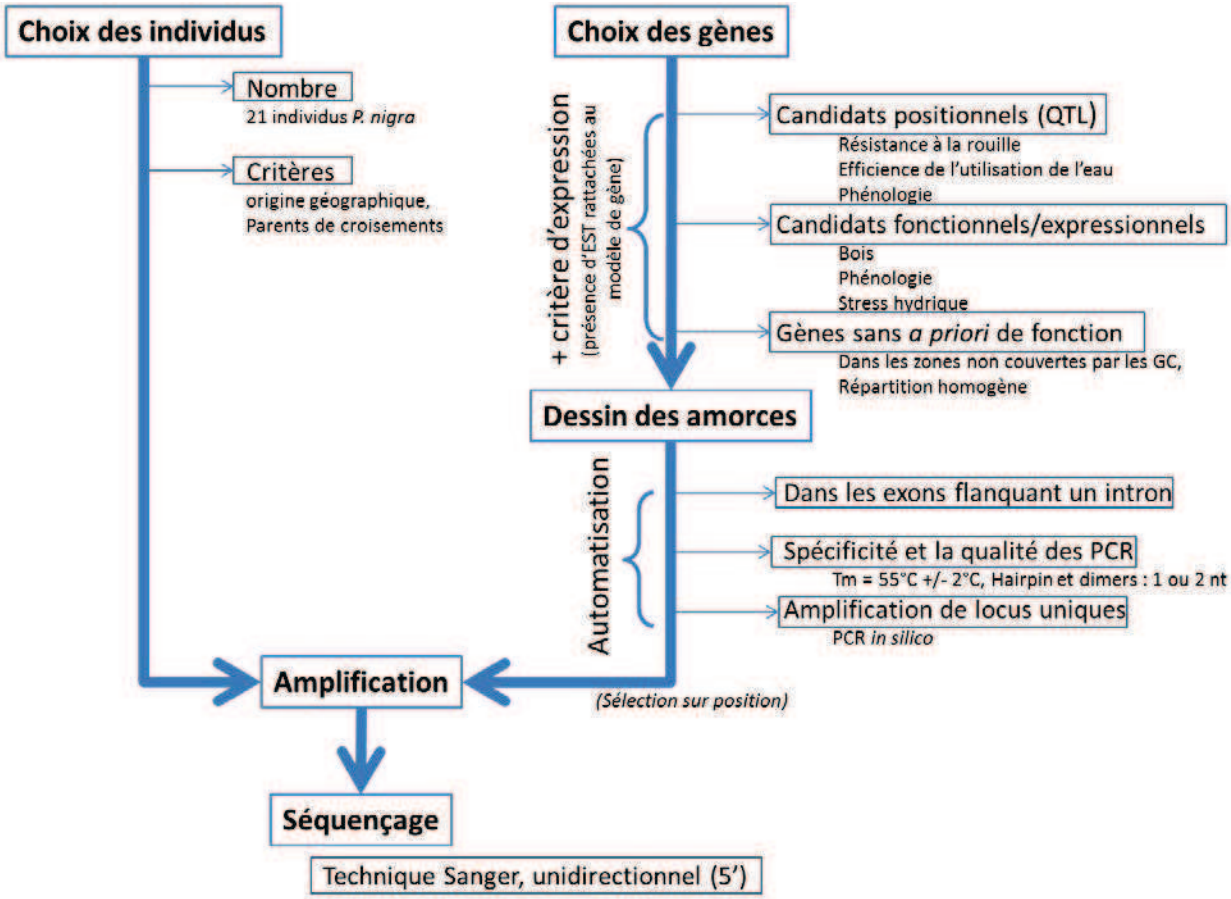


Figure II.2 : Schéma de la stratégie dans la thèse pour l'identification de SNP chez *Populus nigra*.

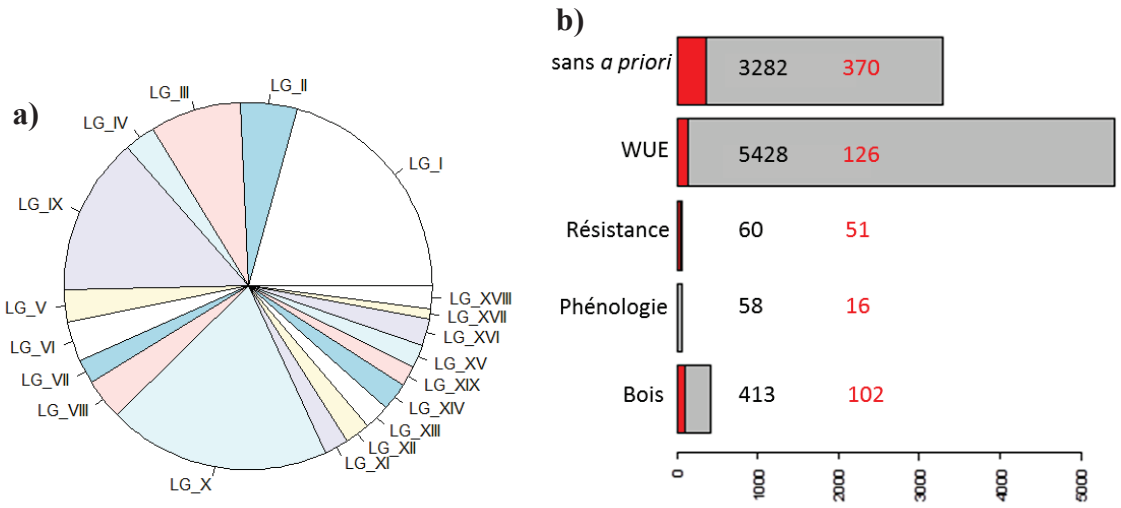


Figure II.3 : Répartition des gènes candidats choisis a) sur tout le génome, et b) par catégorie, avec en noir, les effectifs de départ et en rouge, ceux pour lesquels on a procédé à l'amplification et au séquençage.

- 58 gènes candidats positionnels et fonctionnels associés à la phénologie. Les candidats positionnels correspondent aux QTL détectés pour le débourrement végétatif et l'arrêt de la croissance chez différentes espèces et hybrides de peuplier (Rohde et al. 2011). Les candidats fonctionnels sont impliqués dans le contrôle de la précocité de mise à fleur chez le noyer (C. Breton, comm. pers.)
- 3282 gènes choisis aléatoirement dans les régions génomiques non couvertes par les gènes candidats (les « sans *a priori* »)

Dans le cas des candidats positionnels et sans *a priori*, une validation des modèles de gènes a été effectuée en utilisant comme critère l'existence dans les bases de données d'au moins 5 EST chez le peuplier.

3. Dessin des amorces

Cette liste de gènes a fait l'objet d'un choix d'amorce automatisé développé à Orsay par Vincent Thareau (**Figure II.2**). Les conditions du pipeline ont été très stringentes : Tm = 55°C +/- 2°C et longueur des boucles et des dimères d'amorces < 1 à 2 nucléotides. Les amorces sont essentiellement dessinées pour amplifier les exons avec une préférence d'encadrement d'un intron. En effet, les introns sont connus pour présenter un niveau de polymorphisme plus élevé que les exons. Après le dessin des amorces, un alignement blast a permis d'éliminer les amorces non-spécifiques, puis une PCR *in silico* a été effectuée afin d'éliminer les amorces correspondant à des zones dupliquées. Une sélection supplémentaire sur la base de la position dans le génome a été effectuée pour atteindre 800 fragments de gènes, nombre initialement prévu et contraint par le budget.

4. Conditions de la PCR et du séquençage

Les feuilles des échantillons ont été congelées à -80°C. L'extraction de l'ADN a été effectuée avec le kit Qiagen (DNeasy 96 Plant Kit), le dosage a été effectué par Nanodrop (NanoDrop Technologies, LLC, Wilmington, DEL, USA) et l'ensemble des échantillons a été homogénéisé à 25 ng/µL. Les réactions de PCR et de séquençage ont été réalisées sur la plateforme de l'EPGV (Étude du polymorphisme des génomes végétaux) à Évry. Le volume final de la PCR était de 5 µL, il contenait 2 pmoles de chaque amorce, des dNTP à 0,2 mM, le tampon High Fidelity (Invitrogen), le sulfate de magnésium à 2 mM, et 0,15U de la Taq DNA Polymerase Platinium High Fidelity (Invitrogen). La réaction de la PCR a été faite suivant le programme: 3 min à 94°C, suivi par 30 cycles de (20 s à 94°C, 20 s à 55°C, 2 min à 68°C) et 5 min pour une extension finale à 68°C. Les produits de la PCR ont été purifiés par des colonnes d'agarose et ont été séquencés par la suite en sens « Forward » en utilisant le kit de séquençage Big Dye Terminator

(Applied Biosystems) et une amorce commune SP6 (ATTTAGGTGACACTATAG) intégrée dans les premières réactions d'amplification.

5. Analyse des séquences et de la diversité nucléotidique

L'analyse de séquences demande un savoir-faire en biologie moléculaire et en bio-informatique pour pouvoir analyser l'information portée par l'importante quantité de séquences produites. J'ai été confronté au problème d'automatisation du traitement des données de séquences car il n'existait pas d'outil performant permettant à la fois un traitement automatique des séquences et une prise en compte du cas particulier où les évènements d'insertions/délétions (indel) à l'état hétérozygote décalent le cadre de lecture (cas assez fréquent dans nos séquences). Pour résoudre ce problème, j'ai participé aux essais de versions Béta de deux logiciels bio-informatiques : SeqQual (Lang et al. 2009) et Sniplay (Dereeper et al. 2011). Le développement de ces deux outils n'étant pas suffisamment avancé pour qu'on puisse les utiliser pour l'analyse fine de nos séquences, nous nous sommes alors tournés vers deux autres logiciels existants : polymorfind (http://genhort.angers.inra.fr/projects/polymorfind/) et Genalys (Takahashi et al. 2003). De la comparaison de ces deux logiciels, c'est Genalys qui a été choisi pour le traitement des données, car bien qu'il ne gère pas les indels et que l'analyse soit faite fragment par fragment, la détection des SNP est de très bonne qualité parce qu'on visualise les données de séquences. En revanche, polymorfind gère bien les indels mais il ne permet pas de détecter tous les SNP.

Les séquences au format ABI (format de sortie des séquenceurs) ont donc été alignées en utilisant le logiciel Genalys avec les paramètres par défaut : pour la détection d'un hétérozygote, la hauteur du pic le plus faible correspondant au point analysé doit correspondre au minimum à 20% de la hauteur de l'autre pic (Base calling de 20%). L'annotation des SNP/Indels a été faite manuellement. Le fichier de sortie était un fichier texte tabulé que l'on a transformé en fichier au format fasta. Ces derniers ont été formatés avec le logiciel SeqQual (Lang et al. 2009) pour générer des fichiers d'entrée pour le logiciel Arlequin (Excoffier and Lischer 1995-2009).

L'indice de diversité nucléotidique a été calculé selon la formule : $\pi = \frac{n}{n-1}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} x_i\, x_j d_{ij}$, où n est le nombre de séquences échantillonnées, k est le nombre d'allèles différents, x_i et x_j sont les fréquences respectives des séquences i et j, d_{ij} est le nombre de différences observées entre les séquences i et j (Nei and Li 1979).

Le DL a été estimé avec le r^2. Pour ce calcul, les phases ont été re-construites grâce au logiciel Arlequin avec une approche bayésienne (estimation de la meilleure phase avec

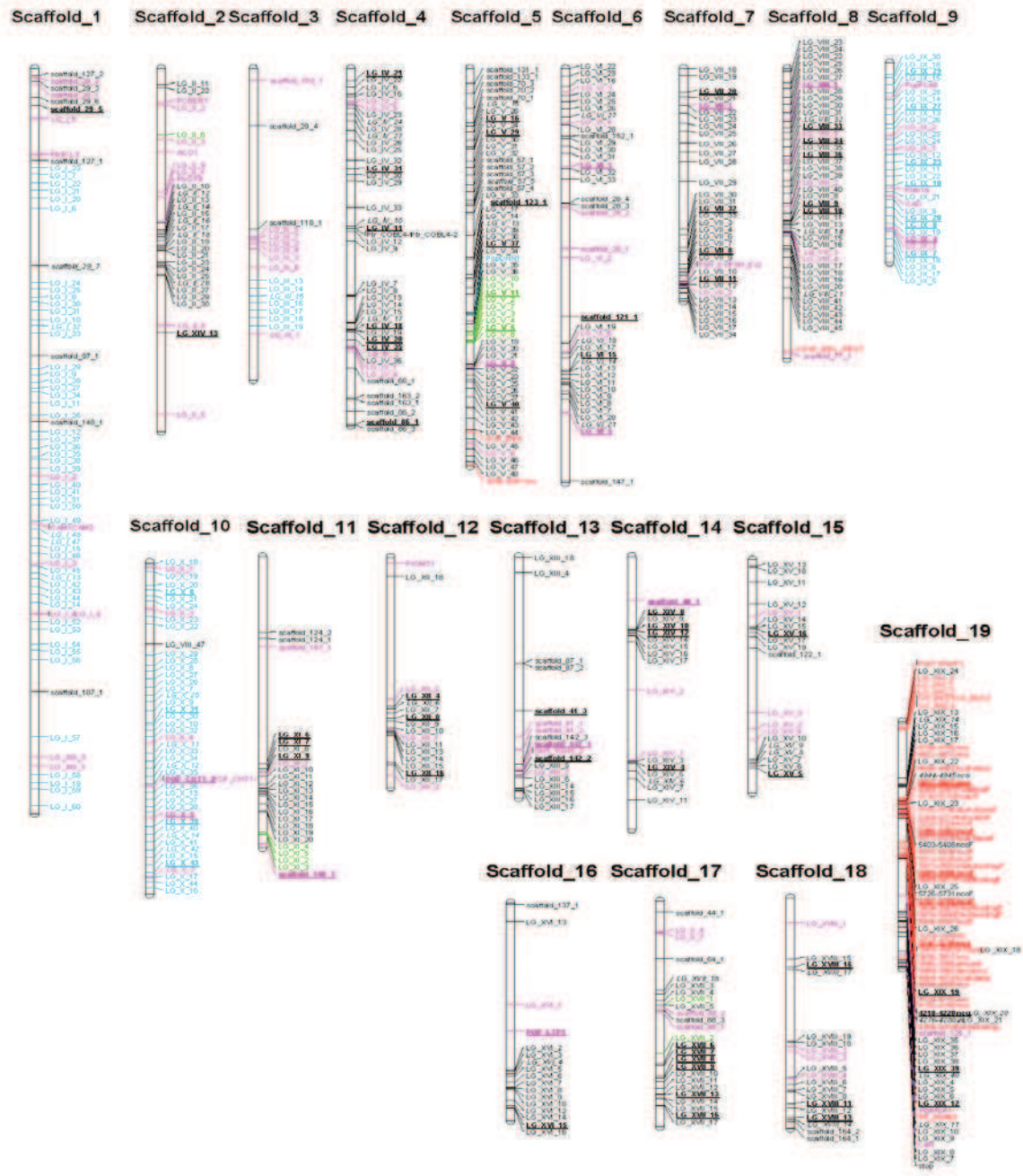


Figure II.4 : Position physique des 665 fragments amplifiés sur la version 2.0 du génome de *P. trichocarpa*. En noir : les gènes sélectionnés sans *a priori*. En rouge: les gènes candidats pour la résistance (expressionnels et positionnels) ; en bleu: les gènes candidats pour l'efficience de l'utilisation de l'eau (WUE, positionnels) et le stress hydrique (fonctionnel) ; en vert: les gènes candidats pour la phénologie (floraison, débourrement, aoûtement. Expressionnels et positionnels). En rose: les gènes candidats pour la qualité du bois (expressionnels et fonctionnels) ; En italique: les gènes qui n'ont pas donné de séquence ou une mauvaise séquence (80 fragments). En gras souligné : les gènes sur lesquels on a développé le panel des 96 SNPs pour la cartographie et le génotypage de la population d'association

l'algorithme ELB). On n'a pris en compte que les probabilités > 0,9 lors de la construction des phases.

Les polymorphismes indel n'ont pas été pris en compte pour le calcul de π et r^2, car dans Arlequin, ils sont considérés comme des SNP et entrainent une sur-estimation de la diversité nucléotidique. On leur a donc attribué le statut de données manquantes.

Un script perl qui parcourt chaque fichier de sortie d'Arlequin de chaque fragment a permis de récupérer dans un même fichier, tous les paramètres calculés pour tous les fragments (indices de diversité, le nombre d'haplotypes ...).

6. Choix de SNP pour la construction d'un outil de génotypage SNPLex® destiné au pédigrée de cartographie génétique

Commercialisée par Applied biosystems, SNPLex était une technologie qui permettait de génotyper 48 SNP en parallèle sur une centaine d'individus. Toutes les régions contenant les SNP ont été alignées (blast) sur le génome de *P. trichocarpa* pour détecter et éliminer les régions dupliquées. Par ailleurs, des contraintes techniques associées à la technique SNPLex® ont réduit le nombre de SNP potentiels : la taille des séquences flanquant le SNP choisi devait dépasser 70 pb et ces séquences flanquantes devaient idéalement être monomorphes. Quand ce n'était pas possible, la distance entre 2 sites polymorphes devait dépasser 15 pb, car si le polymorphisme le plus proche se situait à une distance comprise entre 6 et 15 pb du SNP d'intérêt, il y avait un risque que les sondes ne puissent pas être dessinées. En outre, si le polymorphisme le plus proche se situait à une distance inférieure à 6 pb du SNP d'intérêt, les sondes ne pouvaient pas être dessinées. Les SNP ayant franchi avec succès toutes ces étapes, ont été ensuite évalués par un logiciel de PCR en multiplexage d'Applied Biosystem afin de constituer différentes compositions de 48plex possibles.

C. Résultats

1. Le choix initial et le nombre final de fragments de gènes candidats analysés

Le but initial était d'amplifier 800 fragments de gènes candidats pour un re-séquençage sur 21 individus. À partir de la liste de 9242 gènes choisis sur la base de la séquence de *P. trichocarpa*, le pipeline a réussi à dessiner environ 2400 couples d'amorces encadrant un intron (26,4%). Une sélection sur la base de la position sur le génome a permis de réduire le nombre de couples d'amorces à 800 environ. Au final, on a pu amplifier 665 fragments de gènes répartis sur le génome (**Figures II.3 et II.4**) dont 70% (456) présentaient des séquences de bonne qualité, c'est-à-dire qu'elles étaient lisibles sur environ 500 pb sur la plupart des individus. Un total de

Tableau II.3 : Nombre de gamètes, nombre d'haplotypes et diversité nucléotidique moyens (π) estimés à partir de 456 fragments de gènes dans un panel de 21 *P. nigra*. Ces paramètres ont été calculés au niveau global et au niveau de chaque catégorie de gènes : la formation du bois (62 gènes), la résistance (25 gènes), la phénologie (14 gènes), la WUE (89 gènes) et les « sans *a priori* » (266 gènes répartis aléatoirement dans tout le génome).

	Moyenne		
	Gamètes	Nombre d'haplotypes	π_T
Global	34,9	5,31	0,0028
Bois	36,30	5,08	0,0024
Résistance	31,92	5,64	0,0032
Phénologie	35,08	5,31	0,0025
WUE	35,86	5,55	0,0029
Sans a priori	34,53	5,26	0,0030

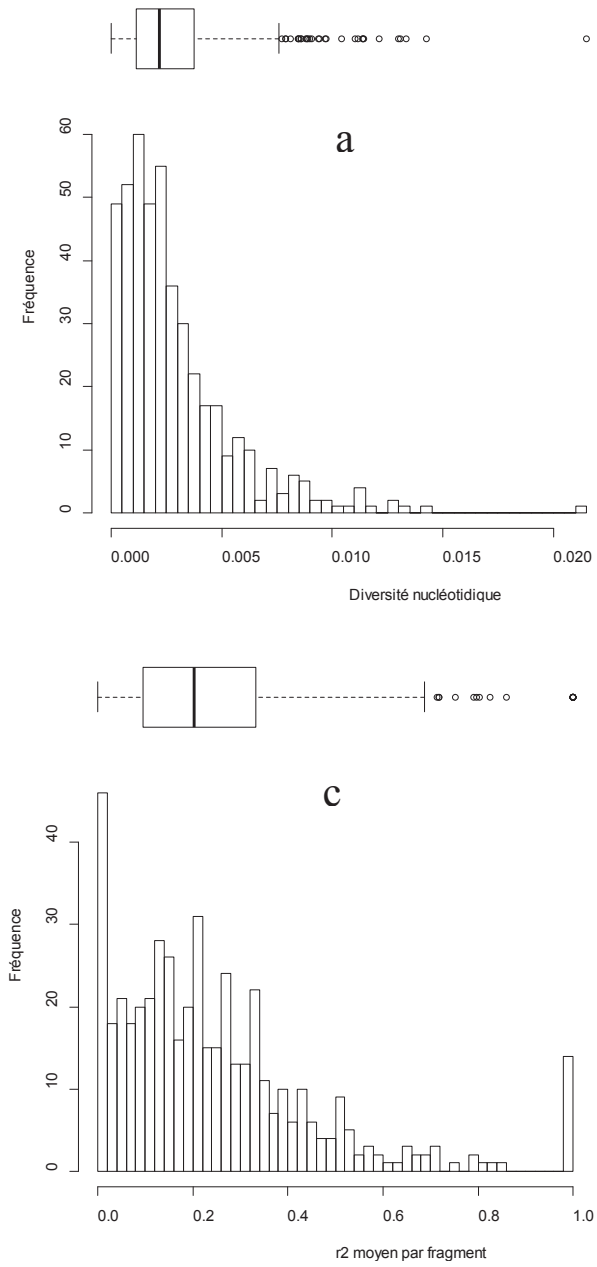


Figure II.5 : Distributions et boîtes à pattes : a) de la diversité nucléotidique et b) du nombre d'haplotypes pour 456 fragments de gènes (34,9 gamètes en moyenne) re-séquencés chez *P. nigra*, et c) du déséquilibre de liaison (r^2) moyen pour 369 fragments de gènes.

Tableau II.4 : Les gènes candidats à valeurs extrêmes extraits de l'analyse de la diversité dans chaque classe (π > 0,0074), avec leur annotation v2.0 du peuplier et l'annotation correspondante d'*A. thaliana* (tair). La classe sans *a priori* est représentée par le candidat le plus extrême (il y en a 20 au total).

Fragment	Modèle de gène v2.0	Defline v2.0	gène ID tair	Defline tair	Classe	π
LG_I_4	POPTR_0001s36690	no poplar v2 defline	AT5G55100.1	SWAP (Suppressor-of-White-Apricot)	Bois	0,008676
LG_VII_2	POPTR_0007s14250	similar to Phosphoenolpyruvate carboxykinase	AT4G37870.1	PCK1 (PHOSPHOENOLPYRUVATE CARBOXYKINASE 1)	Bois	0,008394
scaffold_142_1	POPTR_0013s13470	similar to calcium channel 1	AT4G03560.1	ATTPC1 (TWO-PORE CHANNEL 1)	Bois	0,008466
LG_V_4	POPTR_0005s19110	no poplar v2 defline	AT2G22570.1	nicotinamidase	Phénologie	0,008497
4004-4010hydrolase	POPTR_0019s07550	no poplar v2 defline	NA	no defline	Résistance	0,007791
5222-5224ncoF	POPTR_0019s05180	no poplar v2 defline	AT5G19820.1	embryo defective 2734	Résistance	0,009596
LG_I_48	POPTR_0001s31570	no poplar v2 defline	AT3G25820.1	the monoterpene 1,8-cineole synthase	WUE	0,007778
LG_I_53	POPTR_0001s37380	no poplar v2 defline	AT3G26300.1	CYP71B34; electron carrier	WUE	0,013167
LG_I_56	POPTR_0001s39150	PHD finger family protein	AT3G14740.2	PHD finger family protein	WUE	0,011248
LG_I_59	POPTR_0001s46400	similar to MURUS 4	AT1G30620.2	MUR4 (MURUS 4)	WUE	0,007445
LG_X_21	POPTR_0010s01940	no poplar v2 defline	AT4G27290.1	S-locus lectin protein kinase family protein	WUE	0,011278
LG_X_30	POPTR_0010s09950	no poplar v2 defline	AT2G32150.1	Haloacid dehalogenase-like hydrolase (HAD)	WUE	0,008858
LG_V_16	POPTR_0005s07480	no poplar v2 defline	AT5G10320.1	unknown protein	Sans *à priori*	0,021226

456 fragments de gènes ont ainsi été analysés et se répartissent de la façon suivante dans les différentes catégories de candidats : formation et propriétés du bois (62 gènes), la résistance à la rouille foliaire à *Mlp* (25 gènes), phénologie (14 gènes), WUE (89 gènes) et sans *a priori* (266 gènes).

2. Diversité nucléotidique

Au total, 2609 SNP ont été identifiés dans les 456 fragments de gènes analysés. Les fragments présentant des courts indel représentent 16% des fragments amplifiés. On compte en moyenne 1 SNP toutes les 97 pb.

En moyenne, les séquences d'environ 17 génotypes (34,9 gamètes) ont pu être obtenues (**Tableau II.3**). L'analyse de la diversité nucléotidique et du nombre d'haplotypes a été faite à la fois sur tous les fragments de gènes (niveau global), mais aussi en distinguant les fragments de gènes selon le caractère cible pour lequel ils ont été choisis comme candidats.

Au niveau global, le nombre moyen d'haplotypes est de 5,31, il varie entre 1 et 18 (**Figure II.5**). La diversité nucléotidique moyenne est de 0,0028, elle varie de 0 (10 fragments de gènes) à 0,021. Ces 2 paramètres sont distribués en « L », faisant apparaitre des outliers dans les valeurs élevées. Nous avons donc considéré comme outliers, les valeurs qui étaient plus élevées que les valeurs adjacentes supérieures des boites à pâtes correspondantes (Q3 + 1,5 écart inter-quartile, **Figure II.5**). Ces outliers sont au nombre de 32 (7%) pour la diversité nucléotidique et de 20 (4%) pour le nombre d'haplotypes. Leurs annotations sont décrites pour les fragments de gènes candidats ainsi que pour le fragment de gènes sans *a priori* le plus extrême dans les **tableaux II.4 et II.5**. Tous les modèles de gènes ont des orthologues chez *A. thaliana*. Le gène sans *a priori* ayant la diversité nucléotidique la plus élevée code pour une protéine de fonction inconnue. Trois gènes candidats liés à WUE (POPTR_0001s37380, POPTR_0001s39150, POPTR_0001s46400) ont des valeurs extrêmes pour les deux paramètres, ils possèdent à la fois une diversité nucléotidique élevée et un nombre d'haplotype élevé.

Au niveau des différentes catégories de candidats, la diversité nucléotidique moyenne est légèrement plus élevée pour les candidats associés à la résistance à la rouille foliaire à *Mlp* et à WUE avec des π de 0,0032 et 0,0029 respectivement (**Tableau II.3**), avec toutefois une très forte variabilité intra-classe (**Figure II.6**). Cette diversité est à l'inverse la plus faible en moyenne dans les gènes candidats associés aux propriétés du bois et à la phénologie avec des π de 0,0024 et 0,0025 respectivement, avec là encore une très forte variabilité intra-classe.

Tableau II.5 : Les gènes candidats extrêmes extraits de l'analyse du nombre d'haplotypes dans chaque classe (Nb d'haplotypes >10), avec leur annotation v2.0 du peuplier et l'annotation correspondante *d'A. thaliana* (tair). La classe sans *a priori* est représentée par le candidat le plus extrême (il y en a au total 13 extrêmes).

Fragment	Transcrit v2.0	Defline v2.0	gène ID tair	Defline tair	Classe	Nb. haplotype
LG_XV_2	POPTR_0015s09830	similar to leucine-rich repeat transmembrane protein kinase	AT2G31880.1	leucine-rich repeat transmembrane protein kinase	Bois	12
LG_I_53	POPTR_0001s37380	no poplar v2 defline	AT3G26300.1	CYP71B34; electron carrier	WUE	16
LG_I_56	POPTR_0001s39150	PHD finger family protein	AT3G14740.2	PHD finger family protein	WUE	12
LG_I_59	POPTR_0001s46400	similar to MURUS 4	AT1G30620.2	MUR4 (MURUS 4)	WUE	11
LG_IX_10	POPTR_0009s08770	no poplar v2 defline	AT2G16270.1	unknown protein	WUE	11
LG_X_31	POPTR_0010s09670	no poplar v2 defline	AT3G22520.1	unknown protein	WUE	11
5726-5731ncoF	POPTR_0019s05640	no poplar v2 defline	AT5G27740.1	EMB2775 (EMBRYO DEFECTIVE 2775)	Résistance	12
LG_XVIII_9	POPTR_0025s00560	similar to Pyrophosphate-energized vacuolar membrane proton pump (EC 3.6.1.1)	AT1G15690.1	AVP1; ATPase/ hydrogen-translocating pyrophosphatase	Sans *a priori*	18

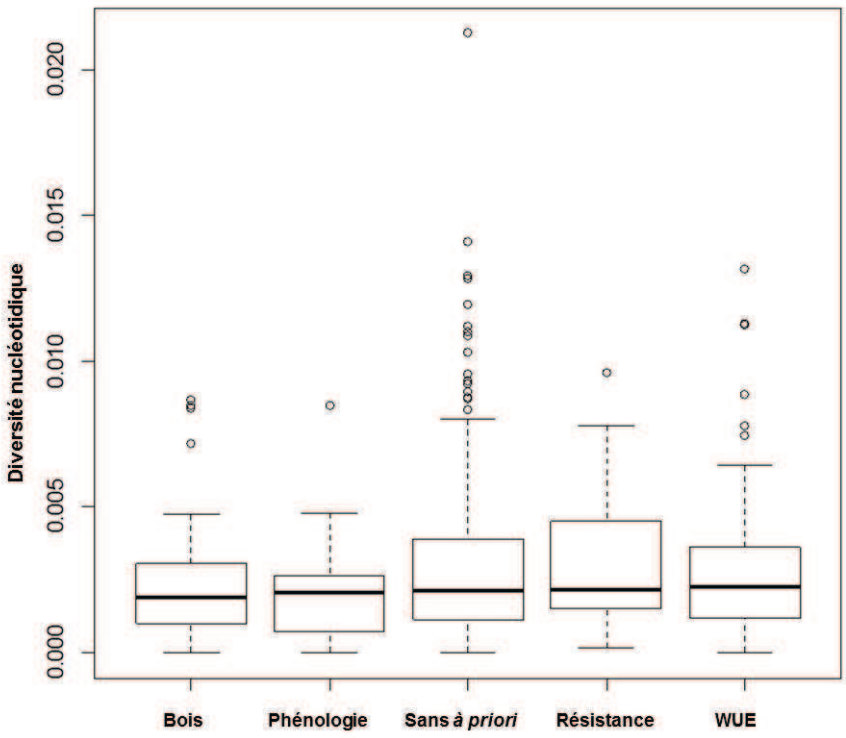


Figure II.6 : Boîte à pattes des valeurs de diversité nucléotidique pour chaque catégorie de gènes (voir effectifs dans le tableau II. 3).

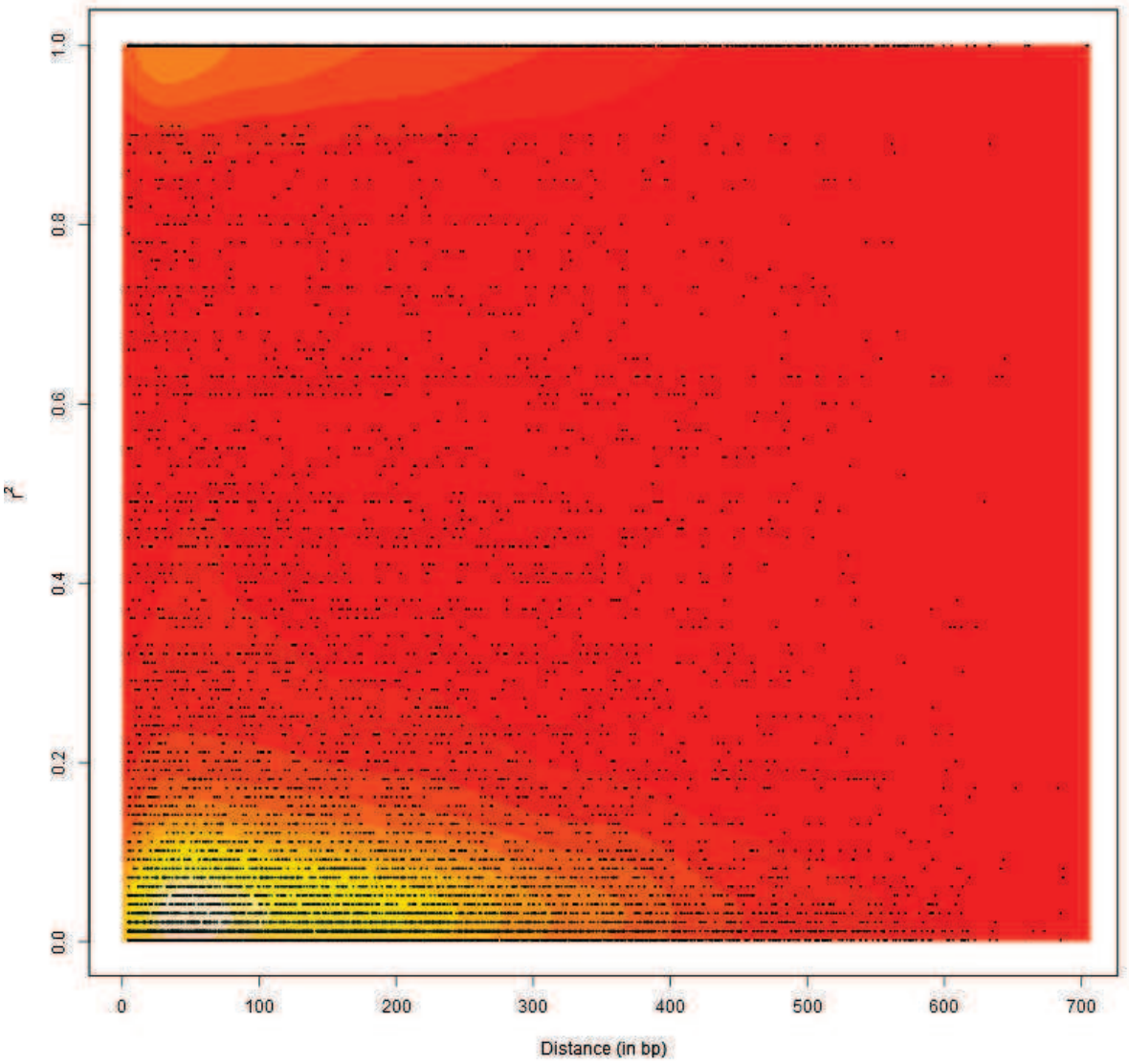


Figure II.7 : Relation entre les valeurs de déséquilibre de liaison r2 et la distance en paire de bases (369 fragments, nombre de points : 11716). Les points sont superposés sur une cartographie de densité de points (plus la couleur tend vers le jaune et le blanc plus la densité en points est forte).

3. Déséquilibre de liaison

Un total de 401 fragments de gènes avaient des haplotypes qui ont pu être reconstruits avec une probabilité supérieure à 0,9 (88%). Parmi eux, 32 fragments avaient seulement un seul SNP. Ainsi, l'estimation du DL a été réalisée à partir de 369 fragments. Le r^2 moyen calculé sur toutes les paires de SNP de tous les fragments de gènes est de 0,31 pour une distance moyenne correspondante de 188 pb. Bien qu'aucune courbe d'ajustement n'ait été tracée, la variation de densité de points permet de dire que le DL chute rapidement à 0,2 au bout de 400 pb (**Figure II.7**).

La distribution des r^2 moyens par fragments de gènes montre elle aussi une forme « L » avec cette fois encore 25 fragments de gènes outliers dont le r^2 moyen était supérieur à 0,68 (**Figure II.5**). Parmi eux, il y avait un gène candidat pour les propriétés du bois, 2 gènes candidats pour la résistance à la rouille foliaire à *Mlp*, 7 gènes candidats associés à WUE et 11 gènes sans *a priori*. Tous ces gènes ont une diversité nucléotidique inférieure à 0,0065 et un nombre d'haplotypes moyen inférieur à 5. La comparaison du r^2 moyen entre les gènes candidats et la catégorie de gènes sans *a priori* ne montre pas de différence entre les deux distributions (**Figure II.8**)

4. Liaisons entre indices de diversité et déséquilibre de liaison

Il existe une liaison positive entre la diversité nucléotidique et le nombre d'haplotypes (**Figure II.9**). La liaison entre le nombre d'haplotypes et le r^2 moyen montre que le r^2 moyen augmente assez logiquement dans les fragments où le nombre d'haplotypes est faible. Enfin, il n'y a pas de liaison évidente entre la diversité nucléotidique et le r^2 moyen par fragment.

5. Relations entre indices de diversité et pourcentage d'introns des fragments

Les relations entre les indices de diversité et le pourcentage d'introns des fragments révèlent des cas particuliers (**Figure II.10**). Par exemple, le nombre d'haplotypes reste faible malgré la présence d'une proportion dépassant les 80% d'introns dans 42 fragments analysés, avec 3 candidats pour les propriétés du bois, 1 candidat pour la phénologie et 9 candidats pour le WUE. De même et de façon remarquable, une diversité nucléotidique élevée a été observée dans un candidat pour les propriétés du bois dont la séquence ne contenait pas d'intron. Il s'agissait du fragment de gène POPTR_0001s36690 qui code pour une SWAP (Suppressor-of-White-Apricot ; **Tableau II.4**). On trouve aussi un r^2 moyen dépassant 0,5 dans des fragments ayant un pourcentage de plus de 80% d'introns. Sept fragments de gènes dont 5 sans *a priori* et 2 candidats pour la phénologie sont dans cette configuration.

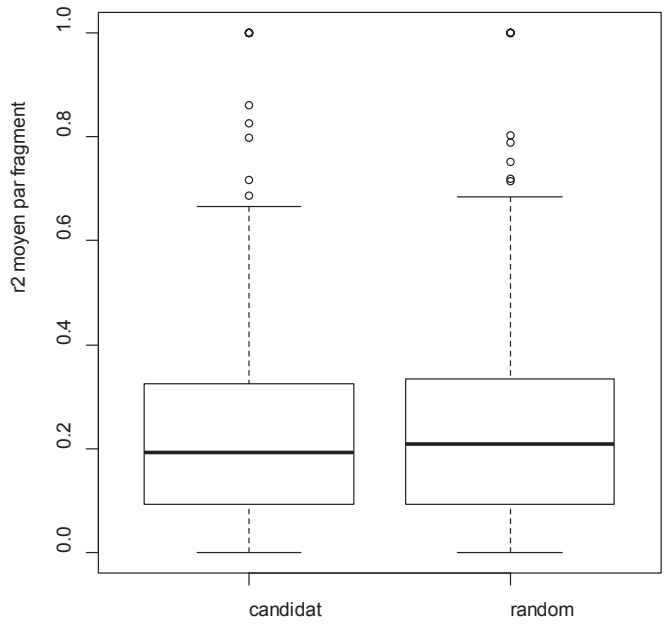


Figure II.8 : Distribution du déséquilibre de liaison (r^2) moyen par fragment pour les gènes candidats et pour les gènes choisis sans *a priori* (random).

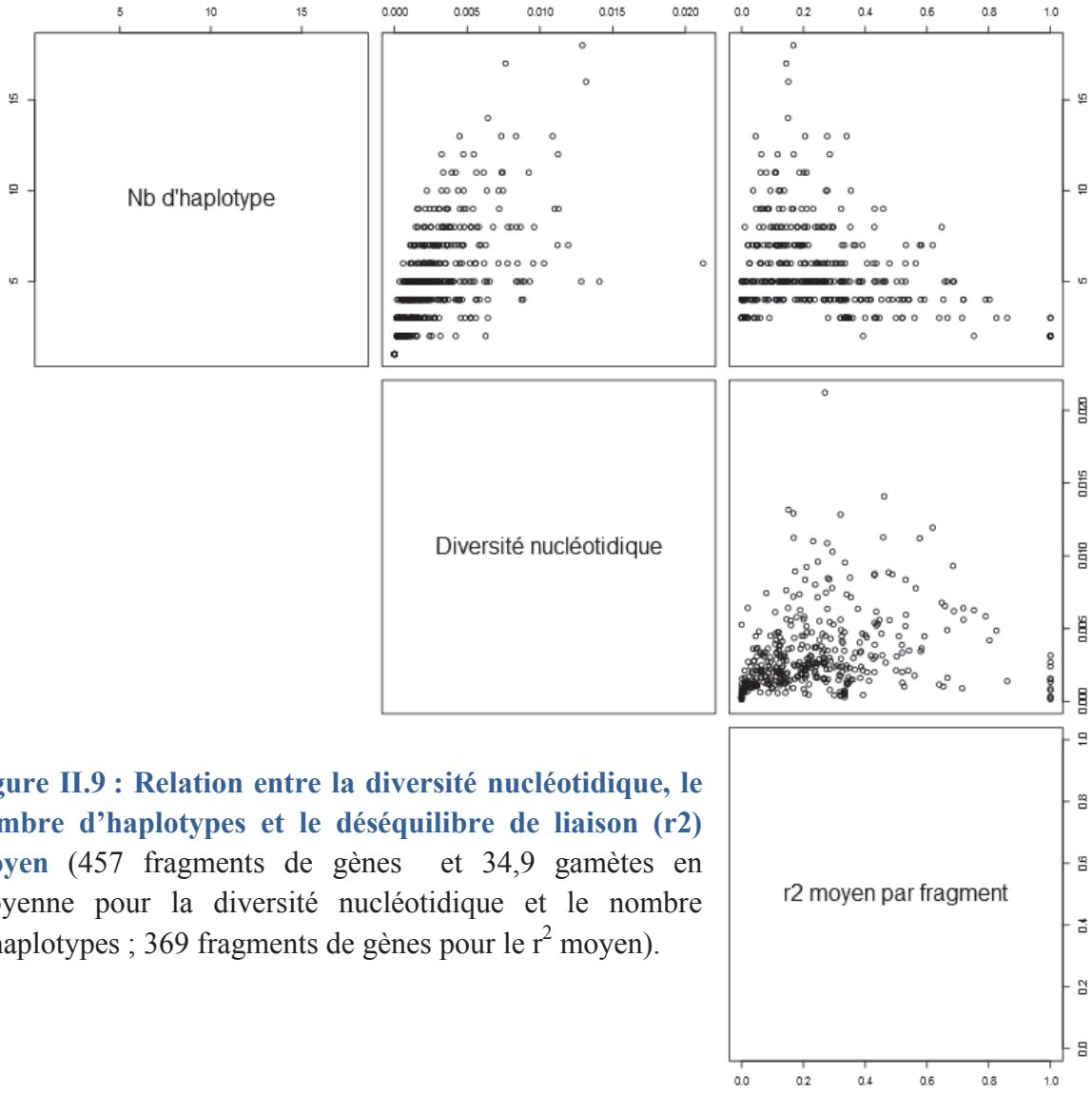


Figure II.9 : Relation entre la diversité nucléotidique, le nombre d'haplotypes et le déséquilibre de liaison (r2) moyen (457 fragments de gènes et 34,9 gamètes en moyenne pour la diversité nucléotidique et le nombre d'haplotypes ; 369 fragments de gènes pour le r^2 moyen).

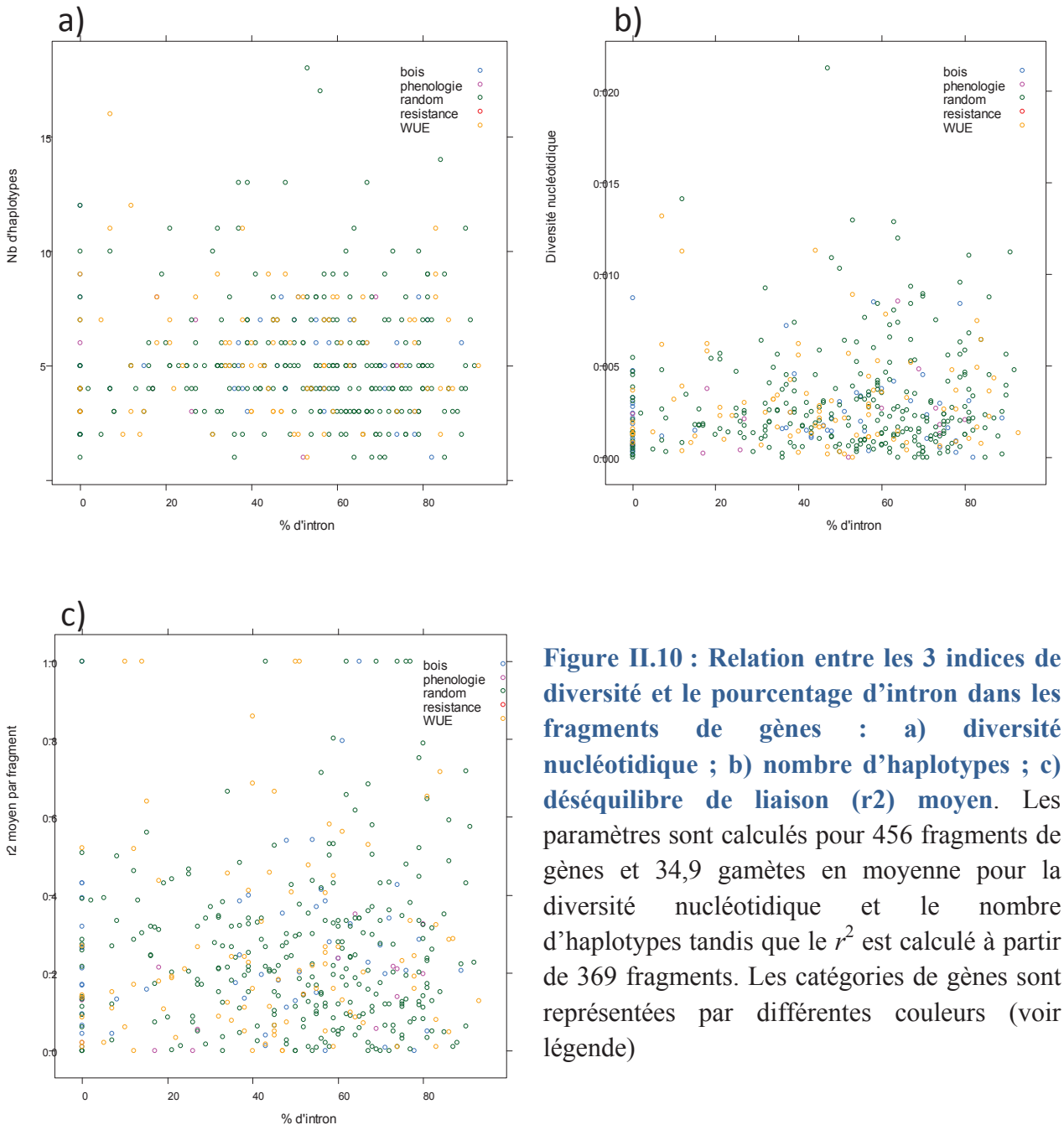


Figure II.10 : Relation entre les 3 indices de diversité et le pourcentage d'intron dans les fragments de gènes : a) diversité nucléotidique ; b) nombre d'haplotypes ; c) déséquilibre de liaison (r2) moyen. Les paramètres sont calculés pour 456 fragments de gènes et 34,9 gamètes en moyenne pour la diversité nucléotidique et le nombre d'haplotypes tandis que le r^2 est calculé à partir de 369 fragments. Les catégories de gènes sont représentées par différentes couleurs (voir légende)

Tableau II.6 : Potentiel de réussite du génotypage de 96 marqueurs SNP (SNPLex) identifiés à partir des deux parents 71072-501 et BDG dans les familles du factoriel de croisement 4x5 *P. nigra*. Les configurations prises en compte sont homozygote chez un parent et ou hétérozygote chez l'autre parent. Les valeurs sont des pourcentages.

	SRZ	**BDG**	71077-308	92501-1
VGN	43	49	33	34
CZB 25	43	46	33	34
71041-3-402	31	55	27	37
71072-501	48	100	48	31

6. Choix de SNP pour la construction d'un outil de génotypage SNPLex® destiné au pédigrée de cartographie génétique

À partir des 665 fragments séquencés, l'analyse des polymorphismes a été entreprise sur les deux parents d'un pédigrée de cartographie génétique (71072-501 et BDG). Parmi les 665 fragments, 111 fragments étaient inexploitables et 276 n'étaient pas polymorphes chez les 2 parents. Ainsi notre choix a pu être effectué sur 278 fragments dans lesquels 631 SNP hétérozygotes chez l'un ou l'autre parent ont été identifiés. Suite à l'alignement (blast) sur la séquence de référence (*P. trichocarpa*) et au filtre lié aux contraintes associées à la technique SNPLex®, il restait 109 SNP disponibles. À partir de ces derniers, le logiciel d'Applied Biosystems a dessiné 2 panels SNPLex de 48 SNP. Ces 96 SNP correspondent à 83 modèles de gènes correspondant à 12 gènes candidats pour les propriétés du bois, 2 gènes candidats pour la phénologie, 6 gènes candidats pour la résistance à la rouille foliaire à *Mlp* et 11 gènes candidats pour le WUE.

En faisant l'exercice de regarder le polymorphisme de ces 96 SNP dans les 8 parents du plan de croisement factoriel, j'ai pu évaluer le taux de SNP également informatif dans chacune des familles correspondantes (**Tableau II.6**). Ce taux varie de 27 à 55% en étant bien évidemment plus élevé dans les lignes et colonnes des 2 parents (71072-501 et BDG) sur lesquels ces SNP ont été sélectionnés.

D. Discussion

Les résultats présentés dans ce chapitre constituent une première étude sur la diversité nucléotidique chez *P. nigra* pour un grand nombre gènes. Le projet a été initié en 2008, avant le début de la thèse, dans le cadre d'un financement AIP séquençage de l'INRA. À l'époque, seules des études de re-séquençage en *Sanger* de quelques gènes candidats chez *P. tremula* (Ingvarsson 2005b, a) et en Ecotilling chez *P. trichocarpa* étaient disponibles (Gilchrist et al. 2006). Depuis, de nombreux travaux concomitants à la présente étude ont été publiés (**Tableau II.1**), et permettent une vision plus globale de la diversité nucléotidique et de son organisation dans le genre *Populus*.

1. Une stratégie de découverte de SNP validée

Le choix de la méthode de séquençage *Sanger*, combiné à une stratégie de conception d'amorces avec des critères stringents et à la disponibilité d'un génome de référence nous a permis d'obtenir un bon taux de réussite au séquençage.

Tableau II.7 : Gènes communs à la présente étude et aux études sur la diversité de gènes candidats chez le genre *Populus* (Ingvarsson 2005b, a; Ingvarsson et al. 2006; Gilchrist et al. 2006; Ingvarsson 2008; Fladung and Buschbom 2009; Chu et al. 2009; Kelleher et al. 2012).

Gène	Defline tair	Nom commun	Catégorie	*P. balsamifera*	*P. trichocarpa*	*P. nigra*	*P. deltoides*	*P. tremula*	*P. tremuloides*	*P. alba*
POPTR_0009s09870	ATCAD4; cinnamyl-alcohol dehydrogenase	CAD-like	Bois		x	x	x	x	x	x
POPTR_0012s08120	no defline		Bois	x						
POPTR_0013s13220	60S ribosomal protein L5 (RPL5B)		Bois	x						
POPTR_0002s06490	GI (GIGANTEA)	GI2	Phénologie	x						
POPTR_0019s14540	no defline (wound-inducible trypsin inhibitors)	TI-3	Résistance	x	x			x		
POPTR_0001s24210	no defline		WUE	x						
POPTR_0001s26060	sugar transporter, putative		WUE	x						
POPTR_0001s36160	G6PD6 (Glucose-6-Phosphate Dehydrogenase 6)		WUE	x						
POPTR_0001s45310	UBP9 (ubiquitin-specific protease 9)		WUE	x						
POPTR_0003s16620	ACR4 (ACT REPEAT 4); amino acid binding		WUE	x						
POPTR_0003s17350	leucine-rich repeat transmembrane protein kinase, putative		WUE	x						
POPTR_0009s06020	WD-40 repeat family protein		WUE	x						
POPTR_0010s00900	ATEBP (ethylene-responsive element binding protein)		WUE	x						
POPTR_0010s23580	no defline		WUE	x						
POPTR_0002s14740	3-phosphoshikimate 1-carboxyvinyltransferase / EPSP synthase		Sans *a priori*	x						
POPTR_0005s17100	CRY1 (CRYPTOCHROME 1)	CRY1.1	Sans *a priori*	x						
POPTR_0005s21640	ARP1 (Arabidopsis ribosomal protein 1)		Sans *a priori*					x		
POPTR_0006s19750	APO2 (accumulation of photosystem one 2)		Sans *a priori*	x						
POPTR_0007s14450	no defline		Sans *a priori*	x						
POPTR_0008s11480	chloroplast nucleoid DNA-binding protein-related		Sans *a priori*	x						
POPTR_0016s11930	PTR1 (peptide transporter 1)		Sans *a priori*	x						
POPTR_0016s12760	pyruvate decarboxylase, putative		Sans *a priori*	x						
POPTR_0017s13070	unknown protein		Sans *a priori*	x						

Pour optimiser la découverte de polymorphisme (objectif premier), notre stratégie était d'amplifier au moins un intron par fragment de gènes. En effet, c'est dans ces régions introniques qu'on attend le plus de polymorphisme comme cela a été décrit chez les plantes (Schmid et al. 2003; Choi et al. 2007; Park et al. 2010). Parfois, cette condition n'a pas pu être respectée pour des raisons techniques. Pour capter le maximum de la variabilité existante dans les gènes, il aurait fallu toutefois séquencer les gènes en entier y compris les 5' et 3'UTR ainsi que les promoteurs (1000 pb en amont du site d'initiation de la transcription). Cette approche de séquençage de gènes complets a été mise en œuvre sur Eucalyptus pour le gène CCR (Thumma et al. 2005). Compte tenu des moyens financiers disponibles, cette approche ne nous aurait pas permis d'analyser autant de fragments de gènes répartis sur le génome et de fait de disposer d'autant de marqueurs d'intérêt pour la cartographie génétique. Les études de diversité moléculaire de gènes candidats fonctionnels et expressionnels montrent que ces deux classes sont peu polymorphes au sein de panels de découverte (Lepoittevin 2009) mais qu'un criblage plus important des variants naturels sur plusieurs centaines d'individus peut s'avérer intéressant pour découvrir des allèles intéressants (Marroni et al. 2011b).

2. Gènes candidats ou sans *a priori* ?

Dans cette étude, deux approches ont été combinées : une approche « gène candidat » et une approche sans *a priori* en essayant de couvrir une grande partie du génome. Il n'y a pas de différence majeure de diversité nucléotidique entre les deux types de gènes. Cela peut être dû au fait que nous n'avons ciblé que des fragments de gènes et non des gènes entiers.

J'ai tenté de comparer la liste de gènes choisis avec les gènes étudiés dans la littérature (**Tableaux II.1** et **II.10**). J'ai rencontré plusieurs problèmes pour faire cette synthèse : d'une part les numéros d'accession pour chaque gène ne sont pas toujours correctement référencés ou parfois ils ne sont même pas fournis, et d'autre part il est assez fastidieux de faire correspondre les modèles de gènes des différentes versions du génome de *P. trichocarpa* (v1.0, v2.2). Finalement, notre liste de gènes contient 23 gènes précédemment étudiés chez d'autres espèces de peuplier (**Tableau II.7**). On trouve un seul gène candidat de résistance Kunitz TI3 (wound-inducible trypsin inhibitor) qui a été analysé chez 3 espèces de peuplier. Pour les gènes candidats pour les propriétés du bois, 9 sont en commun avec deux études de génétique d'association publiées pour la qualité du bois chez *P. trichocarpa* (Wegrzyn et al. 2010) et *P. nigra* (Guerra et al. 2013) et 5 sont en commun avec l'étude de diversité chez *P. tremuloides* (Kelleher et al. 2012). Un gène est particulièrement étudié chez 6 espèces. Il s'agit de la CAD-like (Cinnamyl alcohol dehydrogenase-like) impliqué dans la biosynthèse des lignines. Si on veut comparer les

niveaux de diversité de ces gènes entre les différentes espèces du genre *Populus*, il faut non seulement confirmer que ce sont bien des gènes orthologues mais également que les fragments de gènes coïncident. Cela reste à vérifier, car seules les études sur *P. tremula* portent sur des gènes entiers (Ingvarsson 2005b, a).

3. Des niveaux de diversité et de déséquilibre de liaison très variables.

La diversité nucléotidique moyenne décrite dans cette étude est comparable à celle de la littérature sur différentes espèces de peuplier et de saule. Toutefois ces études ne montrent pas la gamme de variation de cette diversité comme présentée dans ce chapitre. Ici, j'ai pu identifier des outliers positifs à la fois pour la diversité nucléotidique et le nombre d'haplotypes. Il serait intéressant d'étudier de façon plus approfondie ces gènes en les séquençant sur un plus grand nombre d'individus. De manière intéressante, la variation de la diversité nucléotidique n'est globalement pas influencée par la présence d'intron dans les fragments séquencés. Il conviendrait de poursuivre cette analyse en distinguant les SNP synonymes et non synonymes et de comparer les niveaux de diversité correspondants. En effet le pourcentage d'intron a été d'une part estimé sur la version 1.1 de l'annotation, qui a été améliorée depuis ; et d'autre part, il n'est qu'une estimation très mauvaise du nombre potentiel de SNP non synonymes.

Cette étude confirme également le faible niveau de DL chez les peupliers. Toutefois, la comparaison des DL entre espèces et même entre les différentes études n'est pas facile parce que le DL pourrait être affecté par plusieurs facteurs comme la recombinaison, la mutation, la sélection, l'admixture et l'histoire démographique. Ainsi, l'échantillonnage des individus re-séquencés et des régions géniques (ou intergéniques) re-séquencées peuvent influencer les valeurs de r^2. Chez *P. balsamifera,* Olson et al. (2010) montrent que le DL ne décroit pas à l'intérieur des gènes analysés dans un panel de 15 individus couvrant l'aire de répartition naturelle; cependant, Ismail et al. (2012) ont estimé chez cette même espèce un niveau de DL faible qui atteint 0,2 en 250 pb dans un panel de 30 individus échantillonnés dans une plus petite région. Ces résultats illustrent les difficultés de comparaison de ce paramètre entre différentes études même au sein de la même espèce. D'autres problèmes analytiques peuvent aussi intervenir dans les différences observées. Ainsi chez les saules, le DL se maintient à 0,2 au-delà de 1 kb et 4 kb chez *S. schwerinii* et *S .viminalis* respectivement (Berlin et al. 2011). Selon les auteurs, cette différence pourrait s'expliquer par les différences lors de la reconstruction des phases, soulignant l'importance de cette étape.

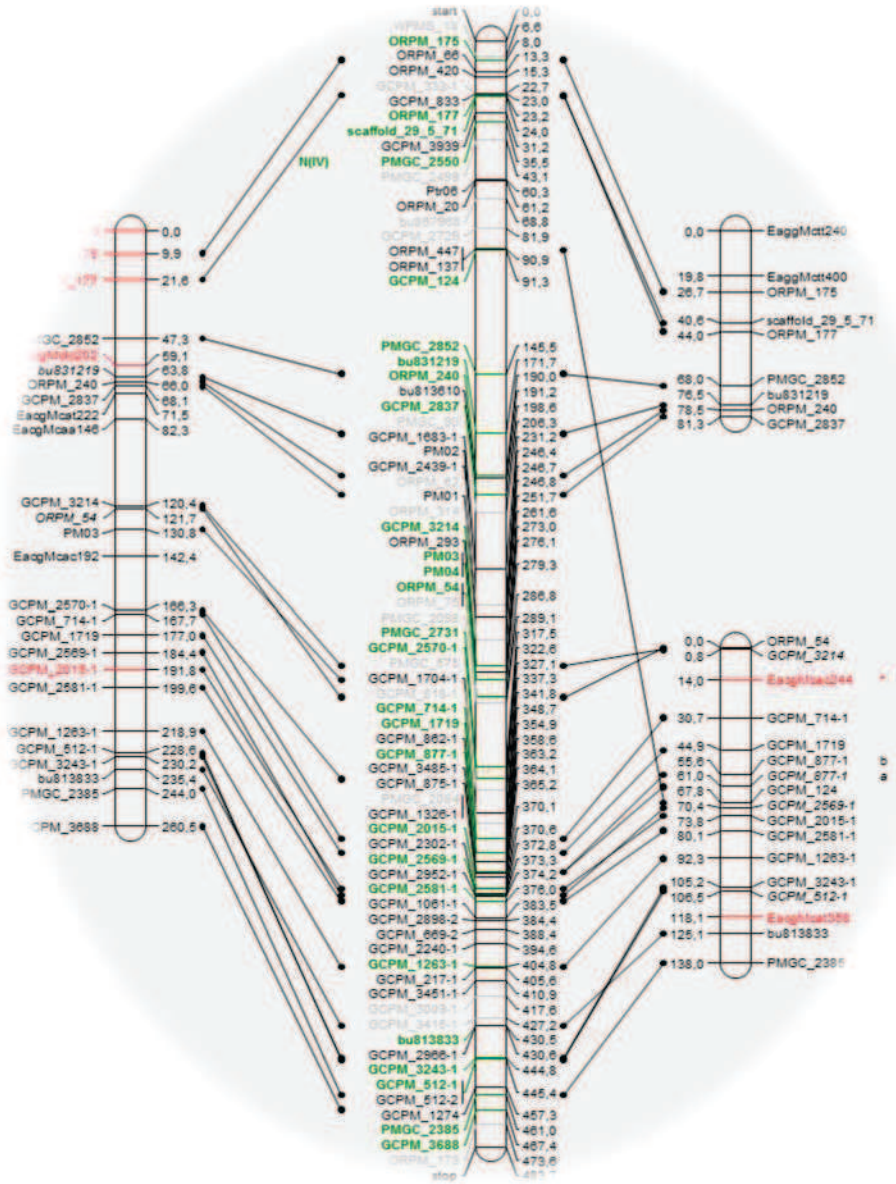
E. Perspectives

Les analyses préliminaires effectuées sur ce jeu de données restent à approfondir :

- Les fragments de gènes ayant des valeurs extrêmes de diversité et de LD seront analysés plus en détail en terme d'annotation (enrichissement en termes GO ; voir chapitre III).
- L'identification des SNP synonymes et non synonymes reste un travail très lourd qui nécessite une automatisation sur un tel jeu de données. Cette automatisation est en cours de développement avec O. Rogier, bio-informaticienne à l'unité. Cette information permettra de calculer non seulement les diversités synonymes et non synonymes, mais également d'effectuer des détections de signature de sélection.
- La disponibilité d'un génome de référence permettra d'analyser la distribution de la diversité nucléotidique le long du génome, et également d'analyser le DL à longue distance sur la base d'un calcul du DL au niveau génotypique.

D'autres données obtenues grâce à ce travail d'identification de SNP n'ont pas été présentées ici : Un total de 415 individus *P. nigra* provenant de 9 populations européennes répartis du sud au nord ont été génotypés à l'aide des 96 SNP utilisés en cartographie génétique. Ce jeu de données pourra venir compléter un article décrivant les résultats présentés dans ce chapitre.

- Développement de nouvelles cartes génétiques chez *P. nigra*



Chapitre III : Cartes génétiques chez *P. nigra*

A. État de l'art

1. Des marqueurs et des cartes génétiques

Historiquement, c'est la génétique quantitative classique basée sur des observations phénotypiques qui a poussé les chercheurs à identifier des marqueurs fiables (héritabilité théoriquement égale à 1) pour pister les caractères d'intérêt et améliorer l'efficacité de la sélection. En 1923, Sax a introduit l'utilisation d'un caractère monogénique mendélien comme marqueur de facteurs génétiques impliqués dans des variations quantitatives du phénotype. Il trouva en effet une association entre la taille du grain du haricot (caractère à variation continue) et la pigmentation des téguments de la graine (caractère mendélien) qu'il attribua à une forte liaison génétique entre le facteur qui contrôle la taille du grain et le locus déterminant sa couleur. Cependant, les possibilités de sélection sur la base de marqueurs phénotypiques à ségrégation mendélienne restaient limitées car ceux-ci ne sont pas nombreux et car leur expression peut être affectée par l'environnement. En outre, des informations importantes pour le généticien restaient encore inaccessibles comme le nombre et l'effet des gènes intervenant effectivement dans l'expression d'un caractère, la localisation de ces gènes sur les chromosomes et leur mode d'action physiologique (place dans les chaînes métaboliques). Grâce au développement d'un très grand nombre de marqueurs moléculaires, notamment ceux basés sur le polymorphisme de l'ADN, un très grand nombre de locus contrôlant l'expression de caractères d'intérêt agronomique ont été identifiés. C'est au travers du suivi de leur ségrégation dans une descendance que l'on peut construire une carte génétique pour non seulement identifier mais aussi localiser sur le génome des locus contrôlant les caractères quantitatifs. L'établissement des cartes génétiques nécessite donc une population (une descendance ou plusieurs descendances) en ségrégation et des marqueurs moléculaires polymorphes (qui ségrégent).

a) *La population en ségrégation*

Le choix des parents d'une descendance dont on veut suivre la ségrégation se base sur une maximisation des chances d'obtenir du polymorphisme moléculaire et de la variabilité phénotypique dans la descendance, ce qui dépend du mode de reproduction de l'espèce. Dans le cas des espèces autogames, ou des espèces allogames chez qui on utilise la consanguinité pour induire la fixation, qui sont essentiellement homozygotes pour tous les locus, et donc peu polymorphes (lignées pures), le choix va porter de préférence sur des lignées divergentes génétiquement ou sur des lignées issues d'espèces différentes. Les espèces à allogamie

préférentielle (pomme de terre, arbres fruitiers et forestiers dont le peuplier), présentent de nombreux locus à l'état hétérozygote. Pour ces espèces, le critère de choix ne peut donc pas porter sur les parents mais se fera plutôt sur le niveau de variabilité génétique dans la descendance.

Une fois les parents choisis, une descendance en ségrégation sera nécessaire pour obtenir un jeu d'événements méiotiques suffisant pour estimer le plus précisément possible les taux de recombinaison entre les marqueurs. Classiquement, le choix du type de pédigrée dépend du niveau de précision recherché mais aussi de contraintes biologiques, techniques et de temps. Par exemple, une descendance F2, issue de l'autofécondation d'un hybride simple F1, peut être obtenue rapidement chez les espèces autogames mais ne permet pas de mesurer les caractères agronomiques de façon précise. Par contre, une population de lignées recombinantes dérivée des individus F2 par 5 à 6 générations d'autofécondation sans sélection (RILs) permet un gain significatif en précision. Pour les espèces forestières, les descendances issues d'une première génération de croisement entre 2 parents (hétérozygotes) sont largement utilisées, et dans le cas d'espèces à multiplication végétative comme les peupliers, il est possible d'atteindre une grande précision dans la prédiction des valeurs génétiques par les phénotypes en particulier pour les caractères peu à moyennement héritables.

La taille de la descendance est l'un des facteurs qui détermine le pouvoir de détection des QTL ainsi que la précision de l'estimation des variables associées (effet, position …). Dupuis et Siegmund (1999) ont montré qu'un minimum de 121 descendants étaient nécessaires pour avoir une puissance de détection de 80% (Dupuis and Siegmund 1999). Mais cette taille optimale de population varie significativement, entre autres, avec le niveau de variabilité et d'héritabilité du caractère, et avec la densité de marquage. Chez les arbres forestiers, les tailles de descendance sont très variables (50 - 994 ; revu par H. Muranty, comm. pers.) et sont le plus souvent un compromis entre le pouvoir de détection de QTL désiré et les moyens de phénotypage et génotypage disponibles !

b) *Les marqueurs moléculaires*

Il existe différents types de marqueurs génétiques : les marqueurs morphologiques, biochimiques et moléculaires (De Vienne 1998). D'après le même auteur, un bon marqueur doit être (i) polymorphe, c'est-à-dire il révèle une variabilité entre les génotypes d'un croisement, (ii) multi-allélique, c'est-à-dire que son polymorphisme peut prendre diverses formes (allèles), (iii) co-dominant, c'est-à-dire que l'hétérozygote présente simultanément les caractéristiques des deux parents homozygotes, il peut donc être distingué de chacun d'eux, (iv) neutre, c'est-à-dire

que les substitutions alléliques au locus marqueur n'ont pas d'autres effets phénotypiques (et donc sélectifs) que ceux de permettre de déterminer son génotype. Sans nécessairement toujours présenter toutes ces caractéristiques, les marqueurs moléculaires, qui sont basés sur le polymorphisme de l'ADN, sont ceux qui présentent les meilleures propriétés et qui de fait sont les plus utilisés pour l'établissement des cartes génétiques. Le polymorphisme de l'ADN peut être classé en trois catégories : le polymorphisme de séquence (SNP), d'indel, et le polymorphisme de nombre d'unités de répétitions dans les régions répétées (Microsatellites ou SSR). Le génotypage permet d'effectuer l'identification et le typage génétique d'un organisme en s'appuyant sur le marquage moléculaire qui permet de déceler les changements au niveau de la séquence d'ADN.

Jones et al. (Jones et al. 2009) ont classé les marqueurs moléculaires en trois générations selon l'historique de leur développement. La première génération de marqueurs moléculaires a été basée sur les enzymes de restriction. Ces enzymes coupent l'ADN en des régions palindromiques bien définies. Dans le cas où le site de restriction contient un changement, l'enzyme de restriction n'opère pas, c'est ainsi qu'on révèle les RFLP (Restriction Fragment Length Polymorphism). Ces marqueurs sont hautement reproductibles cependant les expériences pour les révéler sont fastidieuses (manipulation de la radioactivité pour les hybridations) et nécessitent une quantité d'ADN relativement importante. La deuxième génération de marqueurs moléculaires a été basée sur la PCR (Polymorphism Chain Reaction) ou bien la combinaison de la PCR et des enzymes de restriction. À titre d'exemple, parmi les techniques se basant uniquement sur la PCR, on peut citer les marqueurs RAPD (Random Amlified Polymorphic DNA) où l'amplification aléatoire de fragments d'ADN qui permet de révéler un polymorphisme de fragments amplifiés ; la SCAR (Sequence Characterized Amplification Region) qui consiste à révéler le polymorphisme au sein des fragments clonés issus de la RAPD ; les marqueurs SSR ou encore ISSR (Inter Simple Sequence Repeat) et VNTR (Variable Number Tandem Repeat) qui révèlent le polymorphisme de motifs répétés et les marqueurs STS (Sequence Tagged Site) qui révèlent le polymorphisme de séquences uniques sur le génome très utiles pour l'assemblage de contigs de séquences obtenus après le séquençage de génomes. Parmi les techniques qui combinent PCR et enzyme de restriction, on peut citer la technique CAP (Cleavage Amplification Polymorphism) qui révèle le polymorphisme des sites de restriction après amplification par PCR. A l'inverse, dans le cas des AFLP (Amplified Fragment Length Polymorphism), des S-SAP (Sequence-specific amplification Polymorphism) ou des SAMPL (Selective Amplification of microsatellite Polymorphic loci), le polymorphisme est révélé sur des fragments d'ADN initialement digérés par les enzymes de restriction.

Le séquençage d'un nombre important de gènes chez plusieurs individus voire de plusieurs génomes d'une même espèce a révélé une troisième génération de marqueurs, ce sont les SNP. Ces marqueurs étant extrêmement abondants dans le génome, de nombreuses techniques de génotypage ont alors été développées pour les révéler. Ces techniques varient d'une part selon la méthode utilisée pour cibler l'ADN et d'autre part par la méthode utilisée pour révéler le polymorphisme (Vignal et al. 2002). Des techniques comme la SSCP (Single Strand Conformation Polymorphism), la DGGE (Denaturing Gradient Gel Electrophoresis) et la TGGE (Temperature Gradient Gel Electrophoresis) basées sur la capacité de la renaturation du double brin d'ADN après sa dénaturation permettent de révéler des SNP ou tout changement au sein du double brin analysé. Un SNP peut changer la conformation d'un brin d'ADN ne dépassant pas 300 bases (Vignal et al. 2002).

Le développement des techniques de biologie moléculaire et plus spécialement l'ASO (Allelic-Specific Oligonucleotide), l'OLA (Oligonucleotide Ligation Assay), ou l'extension d'amorces a été derrière un très grand nombre de techniques de génotypage à haut débit. La technique SNPLex (Applied Biosysytems, Californie, USA) combinant ASO et OLA avait le pouvoir de génotyper un multiplex de 48 SNP par expérience. Cette technique a rapidement été dépassée grâce au développement des nanotechnologies, comme la technologie des billes (BeadArray) développée par la firme Illumina (San Diego, USA) (Steemers et al. 2006). Par une extension d'amorces, des fragments spécifiques à un SNP donné sont générés, et les allèles sont révélés après leur hybridation avec des billes contenant des séquences oligonucléotidiques correspondantes à des séquences adresses introduites dans les fragments contenant le SNP. On parle alors de multiplexage à haut débit, la technologie Infinium d'Illumina permettant par exemple de génotyper plus de 100 000 SNP en multiplexage. L'utilisation des enzymes de restriction a toutefois refait surface dans les techniques actuelles à haut débit. Il s'agit des puces de génotypage commercialisées par Affymetrix (Californie, USA) ou encore les DArT (Diversity Arrays Technology, http://www.diversityarrays.com) très utiles pour le génotypage d'espèces dont on ne dispose pas de génomes de référence (Varshney et al. 2010). Actuellement, on parle aussi de possibilité de génotypage par séquençage (GBS ou Genotyping By Sequencing). Cette méthode a été appliquée pour le génotypage du maïs et de l'orge, et représente une méthode de choix pour caractériser les espèces peu explorées jusqu'à maintenant. (Elshire et al. 2011)

Tableau III.1 : Liste des différents logiciels de cartographie génétique avec leur disponibilité, statut et type de pédigrées traités (d'après (Cheema and Dicks 2009)). En souligné, les logiciels et types de pedigree utilisés pour l'analyse dans ce chapitre.

Logiciel	Disponibilité	Statut du logiciel	Pédigrées analysés
AntMap	http://cse.naro.affrc.go.jp/iwatah/antmap/index.html	Version 1.2 réalisée en juin 2006, avec un développement mineur	F2 intercross, F2 backcross, RIL (self), DH
CarthaGène	http://www.inra.fr/mia/T/CarthaGene/	Version 1.0 réalisée en octobre 2004, développement mineur	F2 intercross, F2 backcross, RIL, phase known outbreds
DGMAP	rapidement disparu		Various including F2 backcross
JoinMap	http://www.kyazma.nl/	Version 4 réalisée en juillet 2006 avec une maintenance en aout 2008	BC1, F2 intercross, RILx (self), DH, DH1, DH2, HAP, HAP1, **CP**, BCpxFy, IMxFy
MadMapper	http://cgpdb.ucdavis.edu/XLinkage/MadMapper/	Dernière version en février 2008.	Spécialisé en RILs, peut analyser d'autres familles
MAPMAKER/EXP	http://www.broadinstitute.org/ftp/distribution/software/mapmaker3/	Version 3.0 a été réalisée en décembre 1992. N'est plus en développement	F2 intercross, **F2 backcross**, RIL (self), F3 intercross (self),RIL (sib)
Map Manager QTX	http://www.mapmanager.org/	date inconnue de la dernière version. N'est plus en développement	Advanced intercross, advanced backcross, RILx
MSTMAP	http://www.138.23.191.145/mstmap/	date inconnue de la dernière version	BC1, DH, HAP, RIL
Neighbour Mapping	disponible auprès des auteurs de l'article dont j'ai tiré ce tableau	N'est plus en développement	RIL
RECORD	http://www.plantbreeding.wur.nl/UK/software_record.html	date inconnue de la dernière version. N'est plus en développement	BC1, F2, F3, RIL
THREaD Mapper	http://cbr.jic.ac.uk/dicks/software/threadmapper/index.html	Version 1.0 beta réalisée en juillet 2009.	F2 intercross, F2 backcross, RIL (self), DH

RIL : recombinant inbreed line, **DH** : doubled haploid, **HAP**: Haploid Population, **F2 backcross ou BC1** est la descendance issue du croisement d'un descendant F1 avec l'un des parent homozygote, **BCpxFy** : BC1 répétés suivi d'une autofécondation, p : parents, x : nombre de BC, y : le nombre des autofécondations par Single Seed Descent (SSD) , **IMxFy** : cas général des RIL avec x le nombre de générations issus de croisements aléatoires et y le nombre de générations d'autofécondations par SSD, **CP**: (cross pollinisation) une population issue d'un croisement de deux parents hétérozygotes la phase est inconnue

c) *La construction de cartes génétiques*

Une fois que le jeu de marqueurs a été génotypé sur la descendance choisie, la construction d'une carte génétique peut se résumer en trois grandes étapes : le regroupement des marqueurs, leur ordonnancement et le calcul des distances génétiques entre marqueurs.

Le regroupement des marqueurs dans un groupe de liaison (« grouping ») s'effectue en testant les liaisons deux à deux généralement grâce au calcul du LOD (Logarithm of the odds) score correspondant au rapport des maximums de vraisemblance pour la liaison et l'indépendance des deux marqueurs. Ensuite, plusieurs stratégies existent pour ordonner les marqueurs au sein de ces groupes de liaison (« ordering »). Elles utilisent une fonction objective qui évalue la qualité de l'ordre des marqueurs et classe les différents ordres possibles entre eux. Enfin, il s'agit de calculer des distances génétiques séparant les marqueurs préalablement ordonnés (« spacing ») selon deux fonctions établies par Haldane en 1919 et Kosambi en 1944. La fonction de Haldane prend en compte les enjambements (« crossing overs ») multiples. La distance est donnée par : $dH(cM) = -50 \ln (1-2\theta)$ (θ étant la fréquence de recombinaison). La distance de Kosambi est considérée comme plus proche de la réalité puisqu'il y a des contraintes mécaniques lors de la méiose qui empêchent des enjambements multiples proches (phénomène d'interférence). La distance de Kosambi est donnée par : $dK(cM) = 25 \ln [(1+2\theta) / (1-2\theta)]$. La distance de Haldane reste largement utilisée car le nombre de crossing-over et donc l'interférence (prise en compte par la formule de Kosambi) n'est pas uniforme tout au long des chromosomes.

Il existe différents logiciels permettant d'établir une carte génétique (**Tableau III.1**) (Cheema and Dicks 2009). Je détaillerai ici deux d'entre eux qui sont largement utilisés dans la littérature et que j'ai utilisés au cours de mon travail de thèse. Mapmaker® a été développé pour l'analyse de la première génération des marqueurs moléculaires (RFLP). Il calcule le maximum de vraisemblance de la distance entre deux points ainsi que les LOD scores correspondants. Par exemple, les paramètres utilisés pour la première phase de regroupement dans cette thèse étaient un LOD score de 4 et un taux de recombinaison θ de 0,3. Si le LOD est >4 et que la distance entre paire de marqueurs ne dépasse pas 45,81 cM (Haldane) ou 34,65cM (Kosambi), Mapmaker considère que les marqueurs appartiennent au même groupe de liaison. Pour calculer le taux de recombinaison pour un ordre donné de marqueurs, Mapmaker combine l'algorithme EM (Expectation Maximisation) et les HMM (chaîne de Markov cachée). Pour l'ordre des marqueurs, il établit un test trois points pour tous les triplets de marqueurs liés, et l'ordre établi est perturbé par une fonction « ripple » permettant d'améliorer la fonction objective. Mapmaker fournit alors des solutions alternatives avec les log-vraisemblances correspondantes.

Le logiciel JoinMap® propose quant à lui 4 tests statistiques pour le « grouping » : l'indépendance des LOD score, l'indépendance des p-value, la fréquence de recombinaison et le 'Linkage LOD'. L'ordre des marqueurs et la distance qui les séparent sont établis par deux stratégies : la régression, et le maximum de vraisemblance de Monte-Carlo (Monte Carlo Maximum Likelihood (MCML)). JoinMap propose trois cartes pour chaque groupe de liaison obtenu, allant de la plus probable à la moins probable.

Pour considérer une carte génétique comme saturée, il est d'abord nécessaire que le nombre de groupes de liaison soit identique au nombre de chromosomes du génome haploïde et que chaque marqueur de la carte soit assigné à un groupe de liaison. Mais cela n'est pas suffisant, idéalement, il faut aussi que la distance génétique totale couverte par la carte n'augmente pas lors de l'ajout de nouveaux marqueurs. Cela n'est pas toujours possible, mais il existe des méthodes pour estimer le nombre de marqueurs nécessaires pour saturer la carte en estimant la taille du génome en cM. En supposant une répartition aléatoire des marqueurs, le nombre N de marqueurs en ségrégation nécessaires pour couvrir l'ensemble du génome avec une distance maximale entre marqueurs de X cM dépend de la taille G (cM) du génome haploïde et du risque statistique α accepté : $N \geq log\alpha / \log[1 - \left(\frac{2X}{G}\right)]$. Le nombre d'individus analysés I doit par ailleurs être suffisant pour mettre en évidence un taux de recombinaison θ correspondant à la distance maximale entre paires de marqueurs. En supposant que chaque évènement soit décelable, compte tenu du pédigrée et du type de marqueurs, un effectif minimal peut être calculé : $I \geq log\alpha / \log(1 - \theta)$. La longueur du génome (en cM) peut être évaluée à partir de cartes même partielles. Elle est souvent sous-estimée dans le cas des croisements F2 interspécifiques en raison du plus faible taux de recombinaison de beaucoup d'hybrides (Prat et al. 2006). Les N marqueurs analysés et les K paires de marqueurs correspondantes qui sont liés à une distance maximale X (en cM) sont reliés à la taille G (cM) du génome (Hulbert et al. 1988) : $G = N(N-1)X/K$. Cette estimation peut ensuite être utilisée pour déterminer le nombre de marqueurs nécessaires pour "saturer" un génome avec une densité donnée (Lange and Boehnke 1982). À titre d'exemple, environ 250 marqueurs sont théoriquement requis pour obtenir une couverture de 95% d'un génome de 1500 cM avec un espacement moyen entre les marqueurs inférieur à 20 cM (http://www.pierroton.inra.fr/genetics/cartoqtl/).

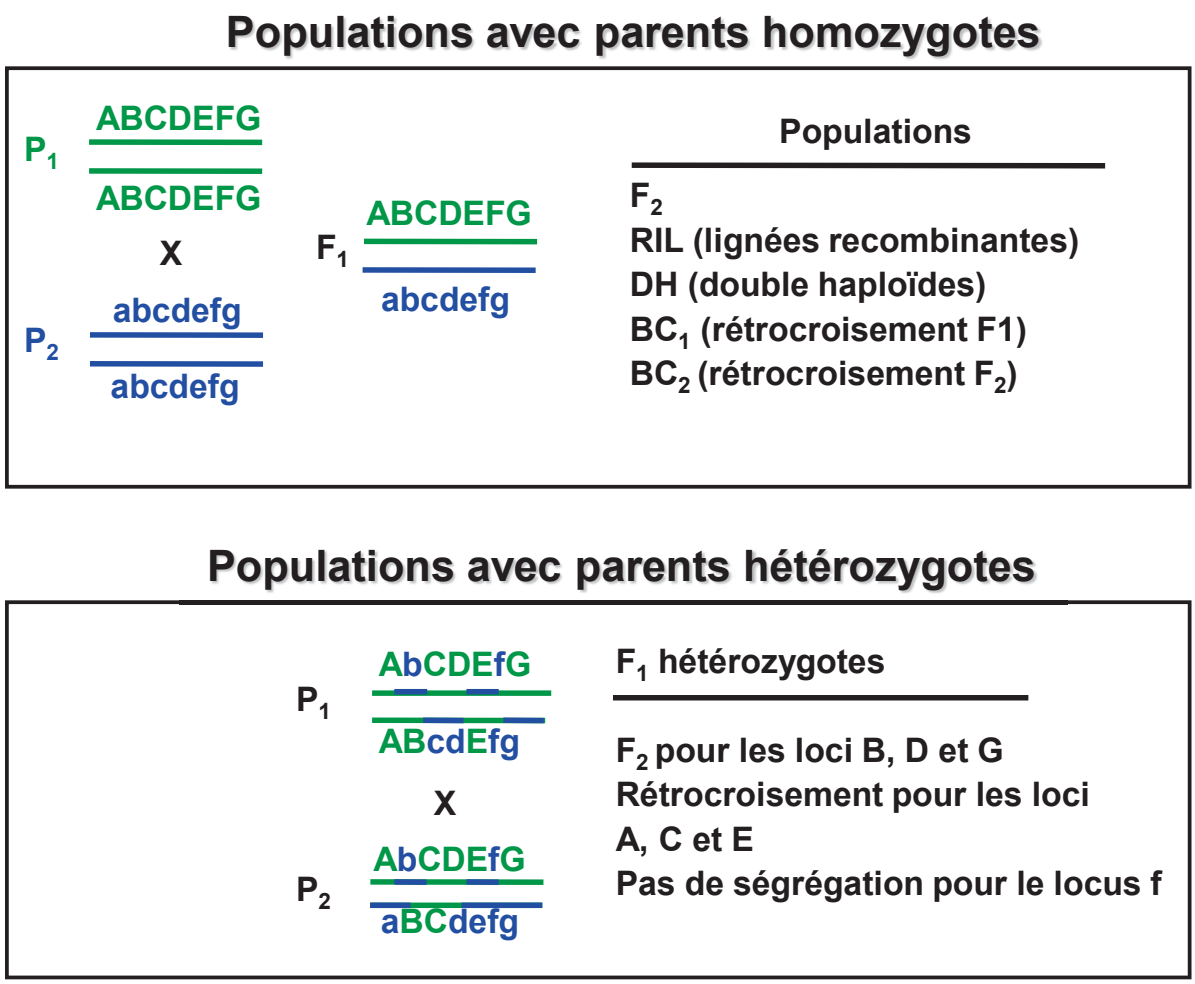


Figure III.1 : Représentation schématique des populations de cartographie génétique. Dans le cas **où les parents sont homozygotes**, la ségrégation des allèles ne va être détectable qu'à partir de la F2. Tous les allèles au sein de la F1 sont en couplage. Dans le cas **de populations issues de parents hétérozygotes**, il y a ségrégation en F1, on peut assimiler la descendance F1 à un rétrocroisement dans le cas des allèles A, C et E, d'où le nom de la stratégie de cartographie génétique (pseudo-testcross). On peut aussi assimiler la descendance F1 à une descendance F2 issue de croisement de F1 hybrides. Les allèles B, D et G illustrent cette configuration. Cependant, La phase entre ces allèles est *a priori* inconnue dans les deux cas.

d) *L'utilisation des cartes génétiques*

Outre leur utilisation majeure pour la détection et la localisation des facteurs génétiques contrôlant les caractères à déterminisme simple ou complexe, les cartes génétiques sont aussi utilisées pour d'autres objectifs, comme le clonage positionnel, la cartographie comparée, ou l'assemblage de génomes. Leurs caractéristiques souhaitées dépendent de l'objectif d'étude. En effet pour la cartographie QTL, un balisage du génome est généralement souhaité. Lors du clonage positionnel de gènes d'intérêt, un nombre très élevé de descendants et le criblage d'un grand nombre de marqueurs sont nécessaires pour aboutir à la cartographie fine du locus d'intérêt (quelques dixièmes de cM). Avec un objectif de cartographie comparée entre espèce étudiée et espèces modèles pour lesquelles on dispose de cartes génétique et physique, il faut un nombre suffisant de marqueurs « ponts », c'est-à-dire des marqueurs qui sont cartographiés dans différentes cartes. De même pour l'assemblage des génomes, on se base sur les marqueurs placés sur une carte génétique très dense et sur leur agencement pour assembler les séquences nucléotidiques.

Enfin, les cartes génétiques peuvent révéler des évènements de sélection (méiotiques ou post-méiotiques) ou de suppression de recombinaison. En effet, ces phénomènes se traduisent souvent par des distorsions de ségrégation regroupées dans une même région chromosomique (Lu et al. 2002 ; Yin et al. 2004b; Torjek et al. 2006) et par des regroupements de marqueurs à un même locus ((Yin et al. 2004b) par exemple).

2. Cartes génétiques développées chez les Salicacées

Chez les arbres, où les populations issues de parents non fixés sont la règle, on a l'avantage d'observer une ségrégation dès la première génération de croisement, mais on obtient un mélange de type de ségrégation (F2, BC1) et on ignore la phase des marqueurs liés (répulsion ou couplage). La différence avec les populations issues de parents homozygotes est illustrée dans la **Figure III.1**. On assiste alors à différents types de ségrégation dans la descendance F1 qui ne peuvent pas être analysés aisément.

Une stratégie fréquemment appliquée est celle du double pseudo-testcross (Grattapaglia and Sederoff 1994). Elle consiste à construire deux cartes à partir du même croisement, en considérant d'une part les configurations de type femelle hétérozygote x mâle homozygote, d'autre part celles de type femelle homozygote x mâle hétérozygote. Dans le premier cas, on n'observera que les ségrégations dues à la méiose du parent femelle, si bien que la carte génétique obtenue ne sera fondée que sur les taux de recombinaison de cet individu. Symétriquement, l'autre configuration

Tableau III.2 : Synthèse des cartes génétiques chez les peupliers. N : nb de descendants dans le pédigrée ; GL : Groupe de liaison ; DPT : double pseudo-testcross.

Nom du pédigrée (espèces)	Publication de la carte génétique	N	Type pedigree/ analyse	Nb et type de marqueurs	Nb de GL	Taille (cM)	Intervalle. moy (cM)	QTL	
								Trait	**Publications**
Famille 331 (TD x TD)	(Bradshaw et al. 1994)	343	F2	344 RFLP, RAPD, STS	35	>1261	6.7	Résistance aux ravageurs, à la sécheresse ; croissance, phénologie, morphologie foliaire	(Newcombe and Bradshaw 1996; Newcombe et al. 1996; Bradshaw and Stettler 1995; Wu and Stettler 1997, 1998; Wu et al. 2002; Ferris et al. 2002)
	(Tschaplinski et al. 2006)	350	F2	1170 SSR, AFLP	22	NA	3.8	Potentiel osmotique	(Tschaplinski et al. 2006)
	(Street et al. 2006)	350	F2	183 SSR, AFLP	22	1,453.1	8	Morphologie foliaire, croissance et biomasse, réponse au CO_2 et à l'ozone	(Rae et al. 2006; Rae et al. 2007; Rae et al. 2008; Rae et al. 2009) (Street et al. 2006; Street et al. 2011)
	(Yin et al. 2010)	310	F2	848 SSR, AFLP, RAPD	20	1927.6	?	Propriétés du bois	(Yin et al. 2010)
Famille 13 (T x D) x D	(Yin et al. 2004b)	180	DPT	? AFLP, SSR	19	2480.7	?	Biomasse aérienne et racinaire	(Wullschleger et al. 2005) (Wu et al. 2010)
Famille 822 (TD x TD)	(Frewen et al. 2000)	346	DPT	318 AFLP, SSR, GC	26 24	2002 1778	13,6 12,3	Phénologie	(Frewen et al. 2000) (Chen et al. 2002)
Famille 52-124 (T x D) x D	(Drost et al. 2009)	418	DPT	167 SSR	19	2970	17.8	Biomasse et qualité du bois	(Novaes et al. 2009)
Famille 87001 (D x N) Famille 87002 (D x T)	(Cervera et al. 2001)	127 105	DPT	448 369 355 339 SSR, AFLP	21 34 23 23	2304 2791 1838 2326	10 12.5 12.4 11.2	Biomasse, phénologie	(Dillen et al. 2009) (Marroni et al. 2011a) (Rohde et al. 2011)
Famille 54 (D x T)	(Jorge et al. 2005)	340	DPT	200 191 AFLP, SSR, RAPD, RFLP	28 36	2803 2740	16.3 17.6	Résistance, phénologie	(Jorge et al. 2005; Rohde et al. 2011)
	(Monclus et al. 2012)	343	DPT	195 208 AFLP, SSR, RAPD, RFLP	24 34	3127 3222	18.1 19.4	Biomasse, efficience de l'utilisation de l'eau	(Monclus et al. 2012)
(N x N)	(Gaudet et al. 2008)	92	DPT	368 317 AFLP, SSR, SNP, sex	20 23	2104 2453	11 13	Phénologie	(Rohde et al. 2011; Fabbrini et al. 2012)
(To x B) x To	(Zhang et al. 2004)	120	DPT	218 144 AFLP	19 19	2683 1956	12.3 13.6	Phénologie, croissance et qualité du bois	(Zhang et al. 2004; Zhang et al. 2006)

permettra de construire la carte génétique du parent mâle. Cette stratégie a été développée initialement chez l'eucalyptus (Grattapaglia and Sederoff 1994).

On peut aussi construire une carte consensus grâce aux marqueurs polymorphes chez les 2 parents qui nous servent de ponts entre les 2 cartes avec 4 classes alléliques, à condition que ces marqueurs soient suffisamment abondants. Cette approche présente l'intérêt de n'avoir qu'une seule carte par croisement mais aussi et surtout de pouvoir identifier des effets de dominance en plus des additivités femelle et mâle lors de la détection de QTL.

Chez le peuplier, la première carte génétique a été construite sur la base de la ségrégation de marqueurs RFLP et STS dans une famille interspécifique F2 issue d'un croisement entre deux hybrides interspécifiques F1 issus eux-mêmes d'un croisement entre un parent *P. deltoides* et un parent *P.trichocarpa* (famille 331 ; **Tableau III.2**). Il faut noter que pour cette première carte, ce n'est pas la stratégie pseudo-testcross qui a été utilisée, mais les auteurs ont fait en sorte de choisir 2 individus F1 comportant un grand nombre marqueurs à l'état hétérozygote ségrégeant aux proportions (1 :2 :1) (Bradshaw et al. 1994). Cette carte a été à la base de plusieurs études de cartographie de QTL sur la résistance aux pathogènes (Newcombe et al. 1996 ; Newcombe and Bradshaw 1996), la résistance à la sécheresse (Street et al. 2006), la phénologie, la croissance et la production de biomasse (Bradshaw and Stettler 1995 ; Wu and Stettler 1997 ; 1998 ; Wu et al. 2002 ; Ferris et al. 2002 ; Rae et al. 2006). Toutefois, elle n'est que très peu ancrée sur le génome de référence (17 STS ancrés sur des groupes de liaison parmi 334 marqueurs de la carte). Une deuxième (Tschaplinski et al. 2006) puis une troisième version de cette carte (Street et al. 2006) ont été ancrées sur ce génome et ont été utilisées pour de nouvelles détections de QTL (Rae et al. 2007; Rae et al. 2008; Rae et al. 2009; Street et al. 2011). Enfin, une quatrième version de cette carte a été publiée en 2010 (Yin et al. 2010) et a servi de support pour la détection de QTL contrôlant les propriétés du bois.

D'après le bilan effectué dans le **Tableau III.2** qui résume les informations disponibles sur les cartes génétiques développées chez le genre *Populus* et utilisé pour la détection de QTL, on constate que peu de pédigrées sont explorés (9 au total). Cela est dû au coût d'investissement très important que représente l'entretien et le phénotypage de familles issues de croisements contrôlés chez les arbres en général et celui du développement de cartes génétiques. Une fois le pédigrée et la carte développés, ils sont valorisés dans la détection de QTL pour de nombreux caractères. Les programmes d'améliorations assez récents peuvent aussi expliquer ce petit nombre de pédigrées de cartographie.

Il existe une grande variabilité dans la taille des cartes (cM). Le type de pédigrée (F1, F2, BC) explique en grande partie cette variabilité. En effet, dans les descendances qui impliquent des parents hybrides interspécifiques de type F2 ou BC, il a été observé un phénomène de suppression de recombinaison dû aux différences génomiques entre espèces (Bradshaw et al. 1994 ; Yin et al. 2004b). Ce phénomène raccourcit donc la longueur des cartes. Par ailleurs, les tailles de pédigrée très différentes pourraient également expliquer cette variabilité.

L'ancrage des cartes entre elles et sur le génome de référence a pu se faire grâce à des marqueurs pour lesquels une information de séquence est disponible. C'est le cas des marqueurs microsatellites (SSR) dont plusieurs milliers ont été développés chez le peuplier (http://www.ornl.gov/sci/ipgc/ssr_resources.htm). Toutefois, il existe relativement peu de ponts entre toutes les cartes génétiques de peuplier. Dans une comparaison incluant 11 cartes génétiques (Cervera et al. 2001; Jorge et al. 2005; Street et al. 2006; Gaudet et al. 2008 ; Paolucci et al. 2010), seulement 5 SSR sont en commun entre toutes les cartes. En prenant les cartes deux à deux, il existe un nombre moyen de 25 marqueurs communs (exclusivement des SSR), ce qui n'est pas suffisant pour orienter les groupes les uns par rapport aux autres (en théorie, il faudrait au moins 2 marqueurs ponts par chromosome c'est-à-dire 2 x 19 = 38 marqueurs). Ceci s'explique par le faible taux de transférabilité des marqueurs microsatellites d'une espèce du genre *Populus* à l'autre, d'autant plus que les espèces sont assez éloignées sur le plan phylogénétique (Tuskan et al. 2004 ; Pakull et al. 2009). Le nombre de ponts entre les cartes et le génome est souvent plus important (V. Jorge, comm. pers.). Cela est dû au fait que pour ponter 2 cartes génétiques, il faut que les marqueurs soient polymorphes chez l'ensemble des parents des pédigrées étudiés, tandis que l'ancrage sur le génome ne nécessite que l'homologie de séquence. Les cartes génétiques comportant de nombreux microsatellites ont permis de révéler qu'une fraction importante de la séquence du génome de *P. trichocarpa* non assemblée v1.1 pouvait être assignée à certains chromosomes (Yin et al. 2004b ; Drost et al. 2009 ; DiFazio et al. 2011) et ont permis la publication d'une version v2.0 de l'assemblage. La convention de numérotation des chromosomes est basée sur une carte génétique publiée en 2001 et qui contenait quelques microsatellites (Cervera et al. 2001). Cette convention est également devenue la règle pour les cartes génétiques de saule.

Un investissement significatif dans l'augmentation de la taille des pédigrées phénotypés et génotypés a aussi permis la cartographie fine de locus impliqués principalement dans la résistance aux rouilles (*M. larici-populina*, *M. x columbiana* ; (Zhang et al. 2001 ; Stirling et al. 2001 ; Bresson et al. 2011). Le locus *Mer* a été cartographié sur le chromosome XIX dans un intervalle de 3,4 cM à l'aide de 525 individus (Zhang et al. 2001). Le locus R_{US} est quant à lui

localisé sur le même chromosome dans un intervalle de 0,8 cM grâce à l'analyse de 1415 individus (Bresson et al. 2011). Ici, l'accroissement de l'effectif permet d'affiner la localisation du locus d'intérêt. Toutefois, malgré un nombre important d'individus (1902), Stirling et al. (2001) obtiennent un intervalle assez large autour du locus *Mcx3* (2,73 cM). C'est le phénomène de suppression de recombinaison dans la région qui pourrait expliquer ce résultat.

Il existe un certain nombre de cartes génétiques qui n'ont pas encore servi pour la localisation de QTL. Ces cartes présentent un intérêt pour la cartographie comparée chez les salicacées parce qu'elles comportent des marqueurs ponts. Il s'agit des cartes effectuées sur les pédigrées (*P. alba* x *P. alba*) (Paolucci et al. 2010); (*P. tremula* x *P. tremuloides*) (Pakull et al. 2009); (*P. fremontii* x *P. angustifolia*) (Woolbright et al. 2008); (*P. adenopoda* x *P. alba*) (Wang et al. 2011) D'autres cartes génétiques ne comportent pas de marqueurs ponts : la carte établie à l'aide de marqueurs AFLP d'un rétrocroisement chez *P. deltoides* (Wu et al. 2000), la carte établie avec des marqueurs RFLP et isoenzymes d'un pédigrée F1 *P. tremuloides* (Liu and Furnier 1993), la carte établie avec des marqueurs RAPD d'un pédigrée F2 (*P. deltoides* x *P. cathayana*) (Su et al. 1998), les deux cartes établies par des marqueurs AFLP d'une famille issue du croisement (*P. tomentosa* x (*P. tomentosa* x *P. bolleana*)) (Zhang et al. 2004) et les deux cartes établies par des marqueurs RAPD avec un pédigrée F1 (*P. alba* x *P. adenopa*) (Yin et al. 2001).

Chez le saule, la majorité des cartes ont été établies dans des croisements intra et interspécifiques entre *S. viminalis* et *S. schwerinii* . Une autre espèce hexaploïde *S. dasyclados* résistante au froid a été introduite dans les schémas de sélection, ce qui a conduit au développement d'une carte génétique pour cette espèce (Ronnberg-Wastljung et al. 2003). Des études ciblées sur certains caractères (sexe, forme de la tige), ont permis la construction de cartes très localisées dans les régions du génome contrôlant ces caractères (Barcaccia et al. 2003 ; Semerikov et al. 2003 ; Lin et al. 2007). Au final, il n'existe qu'un petit nombre de cartes génétiques chez le saule. C'est finalement une espèce très peu explorée par rapport à la variabilité du genre *Salix*. Toutefois, la colinéarité et la synthénie est très conservée avec le génome de *P. trichocarpa* (Hanley et al. 2007 ; Berlin et al. 2010), bien que les marqueurs microsatellites développés chez le genre *Populus* soient extrêmement peu transférables au genre *Salix* (Tuskan et al. 2004). Afin de faire des ponts entre les deux genres, la stratégie a consisté à choisir des gènes sur la séquence du génome de référence *P. trichocarpa*, d'amplifier et de séquencer ces gènes chez les saules pour détecter des SNP qui ségrégent dans des populations de cartographie de ce genre. L'ancrage sur le génome de *P. trichocarpa* se fait sur la base de modèles de gènes (et non de marqueurs).

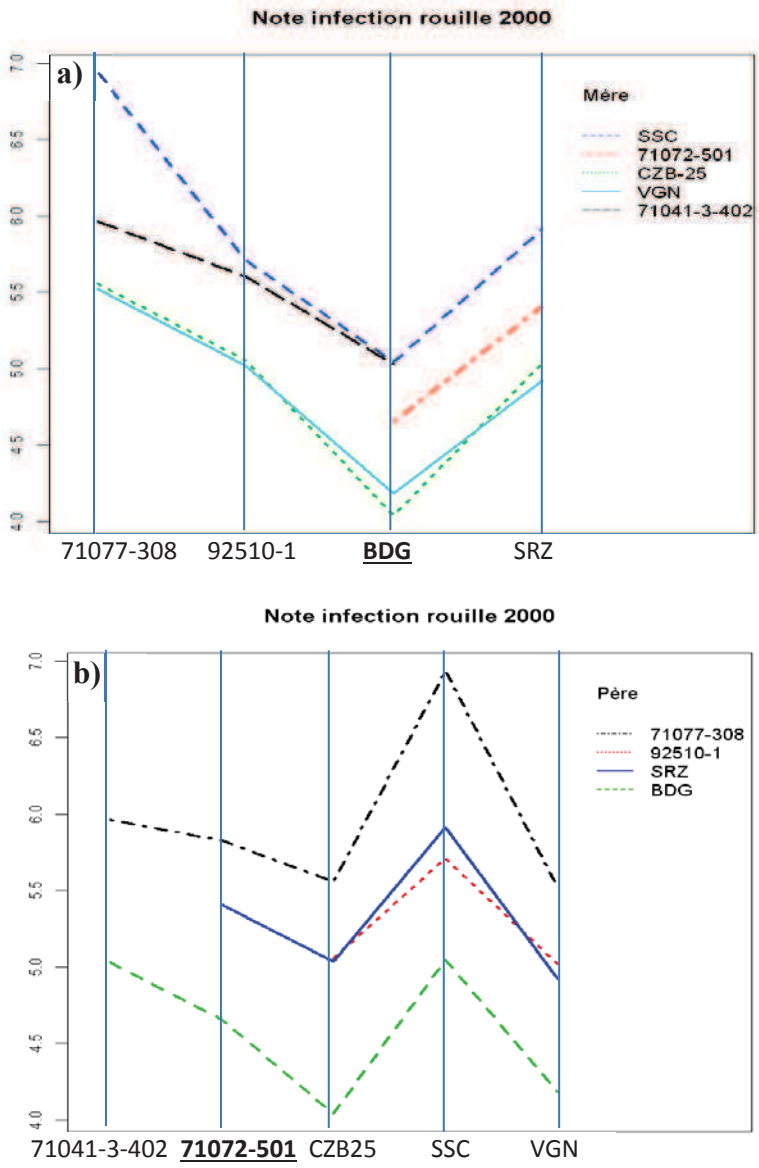


Figure III.2 : Les notes d'infection à la rouille foliaire observées en 2000 pour chaque famille d'un plan factoriel *P. nigra* 5x4 évalué en test de descendance à la pépinière de l'état de Gueméné Penfao. Les courbes représentent les aptitudes générales à la combinaison (évaluées sur les descendances) des a) femelles (71041-3-402, 71072-501, CZB25, SSC et VGN) et des b) mâles (71077-308, 92501-1, BDG et SRZ).

3. Intérêt de la construction de nouvelles cartes génétiques chez *P. nigra*

Chez le peuplier, il existe deux pédigrées impliquant au moins un parent *P. nigra* pour lesquels de la cartographie génétique a été entreprise : un pédigrée intraspécifique (*P. nigra* x *P. nigra*) (Gaudet et al. 2008) et un pédigrée interspécifique (*P. deltoides* x *P. nigra*) (Cervera et al. 2001). Ces cartes nous serviront de référence, cependant l'intérêt de notre nouveau pédigrée de cartographie réside d'une part dans sa taille (324 descendants contre 127 et 92 dans les croisements interspécifique et intraspécifique respectivement), et d'autre part dans sa variabilité phénotypique d'intérêt pour la résistance à la rouille foliaire (voir ci-après et dans le chapitre IV).

B. Outils de conception de cartes génétiques utilisés

1. Famille de cartographie

Une famille F1 a été générée à partir d'un croisement de deux parents *P. nigra*. La femelle '71072-501' a été échantillonnée en 1972 dans une population naturelle originaire du Sud-Est de la France (45°35' N, 05°36' E) le long du Guiers. Le parent mâle est le clone 'Blanc de Garonne' (abréviation 'BDG'). BDG est un cultivar ancien issu de sélection phénotypique réalisée avant les années 1950 au sein d'une population naturelle située en bord de Garonne (la localisation précise de cette population n'est pas disponible). Un total de 324 individus plein-frères constitue cette famille. Cette famille correspond à l'extension d'une famille de 19 individus incluse dans un plan de croisement factoriel 5 x 4 réalisé entre 1990 et 1995 (M. Villar, comm. pers.). L'analyse phénotypique en test de descendance en pépinière de ce plan de croisement avait permis de mettre en évidence une aptitude générale à la combinaison élevée et favorable pour la résistance à *Mlp* (C. Bastien, comm. pers. ; **Figure III.2**). Cette famille est plantée en parc de clones à pieds-mère à la pépinière de l'INRA d'Orléans (47°46'N, 1°52'E, 110m).

2. Optimisation du choix des marqueurs

La stratégie initiale choisie était de combiner des marqueurs microsatellites et marqueurs de type SNP. Les marqueurs SSR présentent l'avantage d'être multi-alléliques et permettront un ancrage de la carte sur d'autres cartes génétiques *Populus* et sur le génome de référence de *P. trichocarpa*. Les marqueurs SNP choisis permettent aussi un ancrage entre cartes (parentales) et avec le génome, mais ils sont aussi plus nombreux et permettent une automatisation du génotypage. Enfin, pour générer rapidement un grand nombre de marqueurs et compléter les cartes génétiques, j'ai aussi choisi la technique AFLP.

Marqueurs SSR

Deux critères de sélection des marqueurs SSR ont été utilisés. Le premier critère était d'utiliser les marqueurs SSR montrant de l'amplification lors de la préparation des cartes *P. nigra* déjà publiées (Cervera et al. 2001 ; Gaudet et al. 2008). Certains d'entre eux monomorphes dans ces pédigrées ne sont pas présents sur ces premières cartes. Cela représente 245 SSR qui amplifient dans l'étude de Gaudet et al. (2008) (M. Gaudet, comm. pers.) et 97 SSR dans l'étude Cervera et al. (2001) (V. Storme, comm. pers.). Le deuxième critère était la présence des marqueurs SSR sur d'autres cartes du genre *Populus*. Cela représente 58, 220 et 217 SSRs amplifiés dans la famille 331 (Street et al. 2006), dans la famille 54 (Jorge et al. 2005) (V. Jorge, comm. pers.) et dans la famille (*P. alba* x *P. alba*) (Paolucci et al. 2010) et (M. Gaudet, comm. pers.) respectivement.

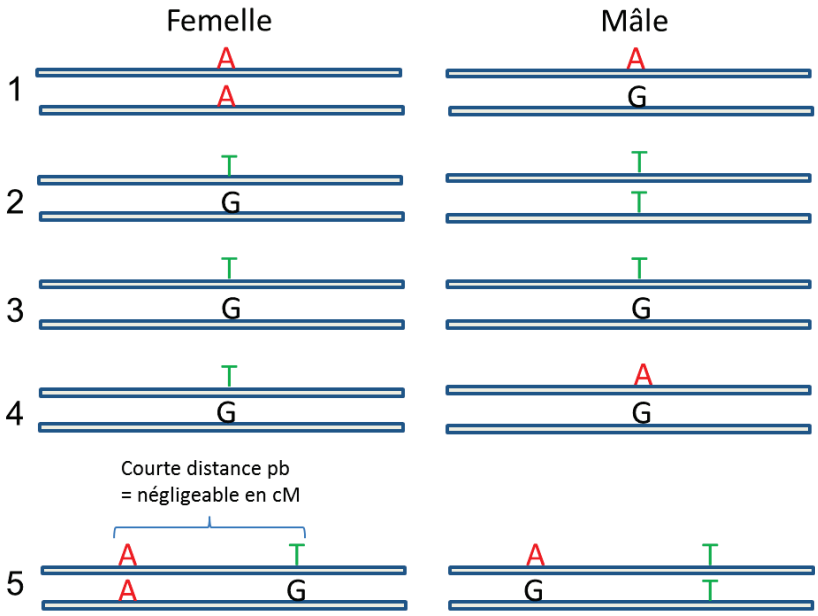
Les séquences des amorces correspondantes sont disponibles sur le site de l'IPGC (International Populus Genome Consortium ; http://www.ornl.gov/sci/ipgc/ssr_resources.htm). Plus de 4 000 SSR sont disponibles dans cette base de données, incluant ceux qui ont été cartographiés mais la plupart n'ont pas été testés et correspondent à des SSR détectés *in silico*. Une autre source de données était les EST-SSR développés à partir des EST par notre laboratoire (Monclus et al. 2012).

Marqueurs SNP

Ils ont été identifiés à partir du reséquençage de 665 fragments choisis dans une liste de gènes candidats dans un panel de découverte de 21 individus *P. nigra*, correspondant à un échantillon représentatif de la diversité naturelle de *P. nigra* en France (voir Chapitre II). Dans notre analyse, 631 SNP hétérozygotes pour l'un ou l'autre des deux parents ont été identifiés.

Un total de 15 marqueurs présents sur les chromosomes I, II et III ont été génotypés à l'aide de la technique PAMSA-M13 (PM). Cette technique est une adaptation de la technique PAMSA (PCR Amplification of Multiple Specific Alleles) développés par Gaudet et al. (2007). J'ai introduit des queues spécifiques M13 dans les amorces et une queue de seulement 3 pb pour différencier les 2 allèles, différence qui peut être distinguée à l'aide d'une séparation sur séquenceur à capillaires (ABI 3100). Dans un deuxième temps, la technique SNPlex (Applied biosystems) a été utilisée et deux panels de 48 SNP ont été élaborés à partir des 631 SNP disponibles.

Afin d'établir des marqueurs SNP ponts, il a été privilégié, quand c'était possible, le choix des SNP homozygotes chez un des parents et hétérozygotes chez l'autre parent et dans le même



	Nb de cas observés : dans les 631 SNP détectés	dans les 96 SNP génotypés
1	392	60
2	239	36
3	0	0
4	0	0
5	-	13 cas

Figure III.3 : Représentation schématique des cas de figure possibles de SNP. Effectifs de ceux retrouvés dans les données de séquence de fragments de gènes chez les deux parents du pédigrée F1 *P. nigra*.

fragment ou un fragment proche physiquement (**Figure III.3**). Les cas 1, 2 et 5 nous intéressaient plus particulièrement. Dans le cas 1, le SNP est hétérozygote uniquement chez le mâle ; dans le cas 2, le SNP est hétérozygote uniquement chez la femelle ; Dans les cas 3 et 4, les deux parents sont hétérozygotes, toutefois, le cas 3 correspond à un SNP moins informatif car l'origine parentale des allèles des descendants hétérozygotes ne peut pas être retracée. Ces derniers SNP n'ont donc pas été comptabilisés dans notre jeu de données. Le cas 4 est un cas rare qui n'a pas été rencontré parmi les 631 SNP détectés chez les parents. C'est pourtant cette configuration qui serait la plus informative pour la cartographie car elle permet d'établir des ponts entre les cartes parentales. Nous avons contourné le problème par une approximation en sélectionnant deux SNP proches sur un fragment, l'un hétérozygote chez la mère et l'autre chez le père (cas 5). La distance maximale rencontrée entre les 2 SNP ne pouvant pas être supérieure à 500 pb (taille moyenne des fragments amplifiés), l'occurrence de crossing-over entre les 2 SNP est très improbable avec l'effectif actuel de la descendance, ils peuvent être considérés comme un seul et même locus.

En respectant les critères génétiques et techniques (voir chapitre II), 109 SNP ont été choisis puis analysés par le logiciel d'Applied Biosystems qui a dessiné 2 panels SNPLex de 48plex et a exclu 13 SNP.

Marqueurs AFLP

Les enzymes de restriction choisies *Eco*RI et *Mse*I sont connues pour révéler un nombre important de bandes et de bandes polymorphes chez le peuplier (V. Guérin, comm. pers.). Ce protocole est une adaptation du protocole initial de développement des marqueurs AFLP (Vos et al. 1995). 5 µL d'ADN de départ est digérée par un mix d'enzymes de restriction (EcoRI 20 U/µL et MseI 10 U/µL) pendant 3 heures à 37 °C puis 10 min à 65°C pour stopper l'action des enzymes de restriction. Une étape de ligation des adaptateurs (AdEco et AdMse) sur les fragments digérés se fait grâce à l'enzyme T4 ADN (400 U/µL) pendant toute la nuit à 14 °C. Ensuite, on effectue une pré-amplification sélective avec les amorces EcoRI-A / MseI-C. Pour l'amplification sélective, j'ai testé 21 combinaisons d'amorces sur les parents et 6 descendants (E-AGG/M-CTG ; E-ACG/M-CAA ; E-ACG/M-CAT ; E-AGG/M-CAG ; E-AGG/M-CTT ; E-ACG/M-CTC ; E-ACG/M-CGA ; E-ACG/M-CAC ; E-ACG/M-CCA ; E-ACG/M-CCT ; E-ACG/M-CG ; E-ACG/M-CGT ; E-AGG/M-CAA ; E-AGG/M-CAC ; E-AGG/M-CAT ; E-AGG/M-CTA ; E-AGG/M-CTC ; E-AGG/M-CGT ; E-AGG/M-CGA ; E-ACG/M-CAG ; E-

ACG/M-CTA). Le génotypage a été effectué au séquenceur à capillaires (ABI 3100) en séparant les produits amplifiés selon leur taille.

3. Élaboration des cartes génétiques

a) *Marqueurs distordus*

La vérification de la ségrégation mendélienne des marqueurs a été réalisée par un test de Chi deux. Lorsque ce test est significatif, les marqueurs ne suivent pas une ségrégation mendélienne et sont dit « distordus ». On a déclaré un marqueur distordu lorsque sa P_{χ^2} était inférieure ou égale à 0,01.

b) *Logiciels de cartographie génétique utilisés*

Dans le cadre de la thèse, j'ai utilisé deux logiciels, Mapmaker® (Lander et al. 2009), très utilisé en cartographie génétique et en particulier chez le peuplier et Joinmap® (Van Ooijen 2006) Joinmap permet d'analyser ensemble des marqueurs ayant des types de ségrégation différents et de construire une carte consensus entre les 2 parents. Il permet d'envisager d'obtenir une carte plus saturée, et présente aussi un intérêt pour la cartographie de QTL (détection des effets de dominance, voir chapitre IV et V).

Dans la stratégie double pseudo-testcross et dans le cadre de l'utilisation du logiciel Mapmaker, on effectue une duplication des données et un recodage inversé de la matrice de génotypes x marqueurs afin de retrouver les liaisons en répulsion. Ceci est illustré dans la **Figure III.4** par un exemple de 2 marqueurs dont le polymorphisme biallélique correspond à la présence ou à l'absence d'une bande amplifiée (type AFLP) qui ségrège dans les proportions 1 :1. La liaison ne peut être détectée que si le marqueur B est codé de façon inversée.

Les paramètres utilisés dans Mapmaker pour le groupage des marqueurs correspondaient à un LOD score de 4 et une fréquence de recombinaison θ de 0,3. Les marqueurs sont regroupés si le LOD est > 4 et si la distance entre marqueurs adjacents ne dépasse pas 34,65cM (Kosambi). Les données analysées par Mapmaker étaient sous forme de matrice 0/1 avec le nom de marqueur et les génotypes correspondants en coupling et en répulsion. J'ai établi deux cartes parentales séparées contenant également les marqueurs distordus.

Dans Joinmap, un travail de codage préalable a été effectué sur les données de génotypage. Il existe un code spécifique pour les allèles paternel et maternel respectivement et le type de ségrégation (1 :1 ; 1 :2 :1 ; 1 :1 :1 :1) ; le nom du type de population est CP. J'ai utilisé le test d'indépendance des LOD scores avec un intervalle de LOD entre 3 et 20. L'ordre des marqueurs et la distance qui les séparent sont établis par régression, avec la fonction de Kosambi pour les

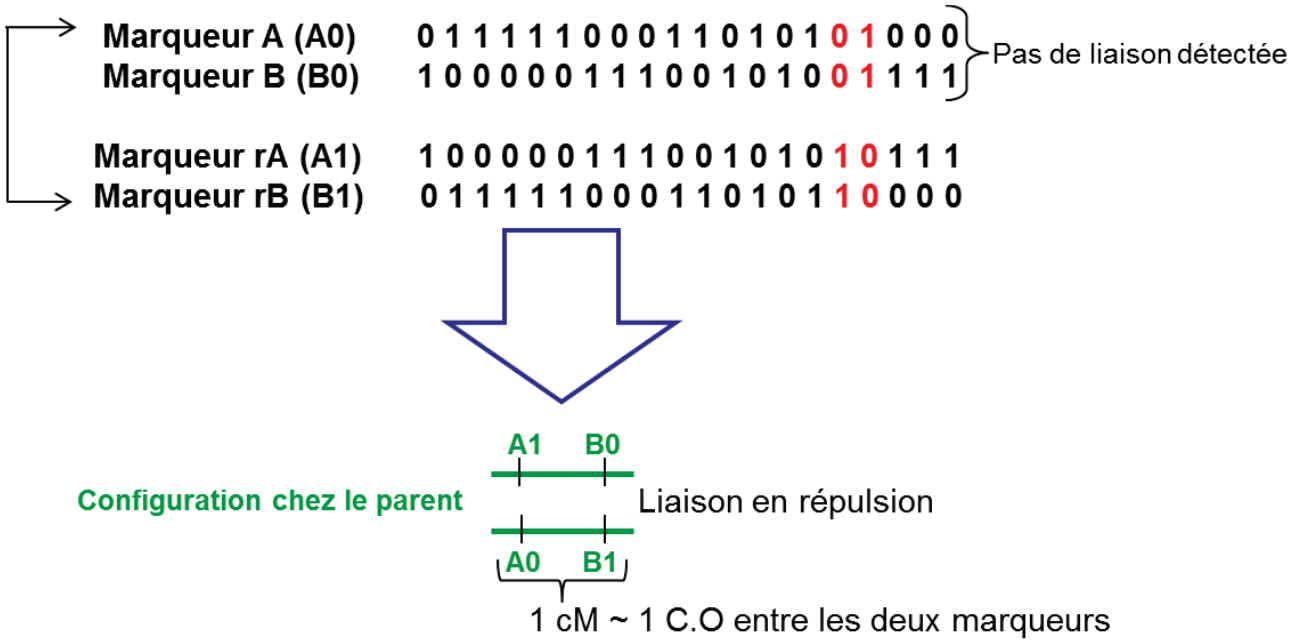


Figure III.4 : Représentation schématique du codage double inversé des marqueurs A et B. Dans ce cas de figure, on ne détecte la liaison que si on a le codage inverse car les locus A et B sont liés en répulsion.

distances. Joinmap propose trois cartes pour chaque groupe de liaison obtenu, allant de la plus probable à la moins probable. J'ai ainsi généré une carte génétique consensus, une carte génétique consensus jalon sans marqueurs distordus et deux cartes jalons parentales sans marqueurs distordus.

Toutes les cartes ainsi construites ont été ancrées sur le génome de *P. trichocarpa* en utilisant les positions des amorces des marqueurs SSR et SNP obtenu par alignement Blast sur la version 2.0 du génome (paramètre Blast par défaut).

c) *Analyse par enrichissement de terme GO*

Grâce à l'ancrage des cartes génétiques sur le génome, il est possible de délimiter des intervalles sur la séquence correspondant aux QTL ou aux zones d'intérêt détectés sur les cartes. Ces intervalles peuvent être relativement grands et comporter plusieurs centaines de gènes dont l'analyse s'avère a priori difficile. Pour analyser de grandes listes de gènes il est possible d'utiliser la classification de la « Gene ontology » et les outils d'enrichissement en termes GO. L'annotation GO du génome *P. trichocarpa* disponible est relativement limitée (18 542 / 40 668 gènes possède une annotation GO dans la version 2.0). En collaboration avec J.C. Leplé, nous avons enrichi cette annotation via l'homologie avec *Arabidopsis thaliana* (Monclus et al. 2012). Les tests d'enrichissements GO ont été réalisés à l'aide du logiciel Ontologizer (Bauer et al. 2008) comme décrit dans Monclus et al. (2012) en utilisant la méthode "Topology-Weighted".

C. Cartes génétiques d'un pédigrée F1 de peuplier noir (Résultats)

1. Criblage de marqueurs et ségrégation

SSR

Parmi les 245 marqueurs SSR amplifiant chez *P. nigra* (M. Gaudet, comm. pers.), 192 (78,4%) ont amplifiés chez les parents étudiés et 113 (46,1%) étaient polymorphes. Parmi les 294 marqueurs SSR provenant exclusivement de la séquence de *P. trichocarpa*, 171 (58,2%) ont amplifié dans les parents étudiés et 66 (22,4%) étaient polymorphes. Ceci souligne l'efficacité du premier critère de choix des marqueurs SSR. Le parent femelle (71072-501) est légèrement moins hétérozygote que le parent mâle (BDG). En effet, sur les 179 marqueurs SSR qui amplifient, 129 sont polymorphes chez 71072-501 contre 149 chez BDG.

Outre la vérification de la ségrégation, le criblage sur les parents et 6 descendants permet d'avoir accès à la fréquence des marqueurs SSR possédant un allèle nul. Ici, 29 marqueurs SSR possède un allèle nul sur les 363 qui amplifient.

Tableau III.3 : Les différents types de marqueurs polymorphes chez le parent femelle et le parent mâle du pédigrée F1 *P. nigra*. Les marqueurs qui ne seront pas liés en carte génétique sont inclus. Le seuil de la distorsion est évalué par un test χ^2. AFLP : Amplified Fragment Length Polymorphism ; SNP : Single Nucleotide Polymorphism ; SSR : Simple Sequence Repeat.

		AFLP		SNP		SSR	
	Marqueurs polymorphes	**Femelle 18**	**Mâle 23**	**Femelle 41**	**Mâle 64**	**Femelle 129**	**Mâle 149**
Marqueurs distordus	0,001<Pχ2 ≤ 0,005	4	3	0	3	3	9
	0,001<Pχ2 ≤ 0,0001	0	0	0	1	5	7
	Pχ2<0,0001	2	1	1	3	14	14
	Total distordus	6	4	1	7	22	30
	% de distorsion	33	17	2,4	11	17	20
	Total/type de marqueurs	10		8		38	
	% de distorsion	22		6,7		21,2	

Tableau III.4 Bilan des caractéristiques des trois cartes génétiques (femelle, mâle et consensus) établies sur un pédigrée F1 *P. nigra*. Les deux cartes génétiques femelle et mâle sont construites par le logiciel Mapmaker. La carte consensus est construite par le logiciel Joinmap. Ce tableau montre un bilan du nombre et du type de marqueurs par carte, la longueur totale de la carte, la distance moyenne entre paire de marqueurs voisins et le nombre de groupes de liaison (GL) par carte dont les doublets (GL avec deux marqueurs seulement) et les triplets (GL avec trois marqueurs seulement). Les marqueurs ponts sont à la fois sur la carte femelle et la carte mâle et sont déjà comptabilisés dans chacune des cartes parentales.

Cartes	**AFLP**	**SNP**	**SSR**	**Non liés**	**Total liés**	**Longueur (cM)**	**Distance moyenne entre marqueurs voisins (cM)**	**Nombre de GLs**
Femelle	18	41	129	14	174	1888,2	12,84	25 (6 doublets)
Ponts		13	99	3	109			
Mâle	23	64	149	12	224	2406,1	12,43	22 (3 doublets et 1 triplet)
Consensus	41	105	179	15	297	2188,7	10,27	24 (2 doublets)

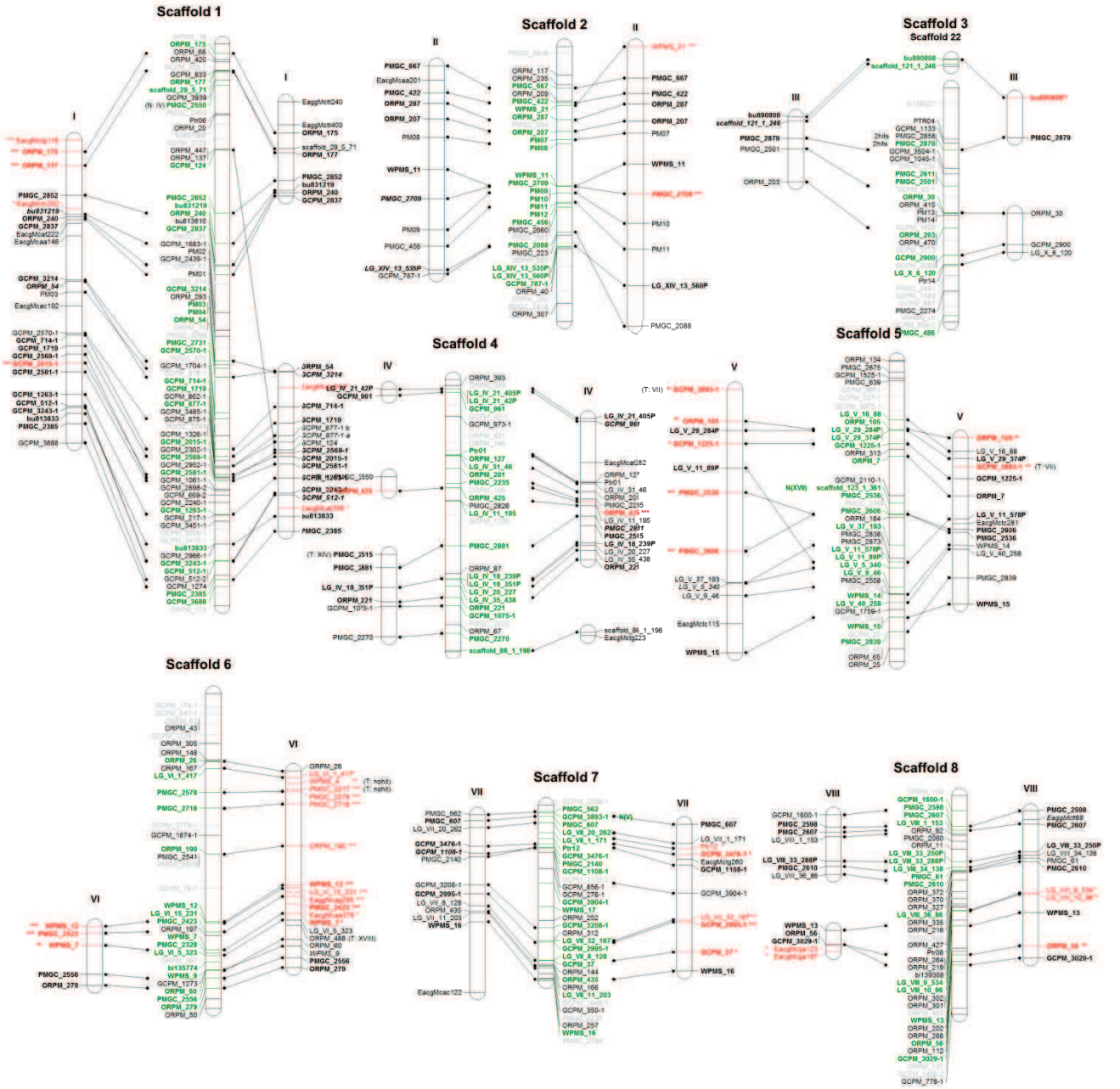


Figure III.5 : Alignement sur le génome de *P. trichocarpa* Nisqually-1 v2.0 des cartes génétiques parentales basées sur l'analyse de la ségrégation de marqueurs SSR, SNP et AFLP dans la descendance du croisement entre *P. nigra* 71072-501 x BDG (319 individus F1). Les cartes génétiques ont été construites avec le logiciel Mapmaker et les paramètres décrits dans le chapitre IV. La position de maqueurs sur la séquence du génome est déterminée par alignement BLAST des amorces SSR et des sondes SNP. Pour chaque chromosome, de gauche à droite, la carte génétique maternelle (71072-501), la position des marqueurs sur le génome et la carte génétique paternelle (BDG). En gras, les marqueurs ponts entre cartes génétiques ; en rouge, les marqueurs distordus, **: $0.01 < \chi2$ P-value ≤ 0.05, **: $0.001 < \chi2$ P-value ≤ 0.01, *** $0.001 \leq \chi2$ P-value; En italique, marqueurs accessoires ; Les discordances (non synthénie) entre cartes génétiques et génomes sont indiquées à côté du marqueur (par ex. : (N : IV) le marqueur est localisé sur le GL IV des cartes génétiques) ; Tous les marqueurs testés dans la thèse ont été positionnés sur la séquence du génome : en gris, les marqueurs qui n'amplifient pas ; en noir, les marqueurs monomorphes ; en vert gras, les marqueurs cartographiés (polymorphes).

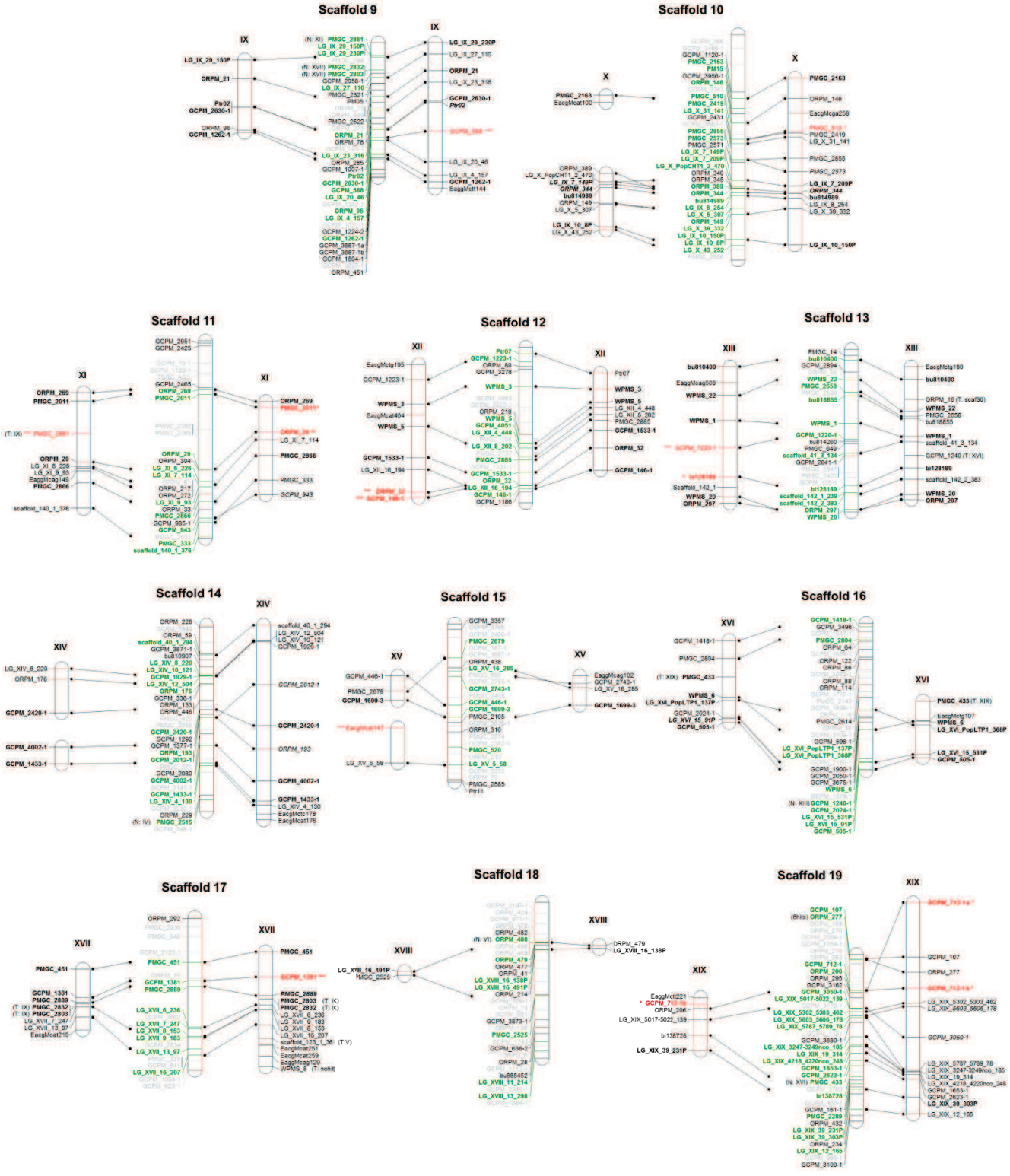


Figure III.5: (suite)

SNP

Le taux de succès (100%) du génotypage par SNPlex est important si on ne tient pas compte de tout le criblage sur les séquences effectué préalablement, soulignant la robustesse de la méthode de génotypage SNPlex. Si on reprend l'ensemble des fragments séquencés (665 fragments), le mâle BDG est plus polymorphe que la femelle 71072-501 : sur les 631 SNP détectés chez les parents, 392 (62,1%) étaient polymorphes chez BDG et 239 (37,9%) chez 71072-501.

AFLP

J'ai retenu 8 couples d'amorces pour les AFLP, celles qui révélaient au moins 4 fragments polymorphes (E-AGG/M-CTG ; E-ACG/M-CAA ; E-ACG/MCAT ; E-AGG/M-CAG ; E-AGG/M-CTT ; E-ACG/M-CTC ; E-ACG/M-CGA ; E-ACG/M-CAC). Ces combinaisons devaient en théorie amplifier un total de 101 fragments polymorphes (sur la base de 6 individus de la phase de test) ; mais l'amplification sur l'ensemble des descendants a donné un total de 41 fragments polymorphes. Vingt-trois marqueurs AFLP proviennent du mâle BDG et 18 marqueurs AFLP proviennent de la femelle 71072-501.

La distorsion des marqueurs

Le pourcentage de marqueurs distordus est sensiblement le même pour les 2 parents (17% et 16% respectivement pour la femelle et le mâle (**Tableau III.3**). Ce pourcentage est différent selon les catégories de marqueurs. Ainsi, il y a moins de marqueurs SNP distordus (6,7%) que de marqueurs AFLP et SSR distordus (22 et 21,7%).

2. Construction des cartes génétiques parentales et consensus

Trois cartes génétiques ont été construites. Deux cartes parentales et une carte consensus. Leurs caractéristiques sont résumées dans le **Tableau III.4**. La distance moyenne entre paires de marqueurs voisins est de 12,84 cM, 12,43 cM et 10,27 cM pour les cartes femelle, mâle et consensus respectivement. La construction de la carte consensus a réduit la distance génétique totale mais n'a pas réduit le nombre de groupes de liaison (24 groupes de liaison) qui reste quant à lui plus élevé que celui obtenu pour la carte mâle (22 groupes de liaison). La carte femelle est moins longue que celle du mâle, ce qui est consistant avec leurs niveaux d'hétérozygotie estimés lors de l'analyse de leurs séquences. Le groupe de liaison XVIII est le groupe de liaison le plus court, il est représenté par un doublet chez les deux parents. Une région distordue de 12 marqueurs (2 AFLPs, 8 SSRs et 2 SNP) est repérée sur le groupe de liaison VI de la carte mâle.

Tableau III.5 : Discordances entre les cartes établies dans le pédigrée F1 *P. nigra*, les cartes génétiques publiées et construites sur la base de croisements impliquant au moins un parent *P. nigra* (Gaudet et al. 2008 et Cervera et al. 2001), et les différentes versions d'assemblage du génome *P. trichocarpa* (Tuskan et al. 2006 ; www.phytozome.net).

	Cartes génétiques			Génome		
Nom des marqueurs	**Elmalki et al. soumis**	**Gaudet et al. (2008)**	**Cervera et al. (2001)**	**Nisqually-1 V1.0**	**Nisqually-1 V2.0**	**Nisqually-1 V3.0**
ORPM_269	XI	VIII, XI	_	XI	scaffold 11	scaffold_1984&Chr11&Chr08
PMGC_451	XVII	XVII	C	scaffold_21039	scaffold 17	Chr17
PMGC_2550	IV	I	-	I	scaffold 1	scaffold_670&Chr01
WPMS_6	XVI	XVI	XII	no hit	scaffold 16	no hit
WPMS_4	VI	VI	-	VI	no hit	no hit
ORPM_16	XIII	XIII[a]	-	XIII	scaffold 30	no hit
PMGC_2832	XVII	-	-	XVII	scaffold 9	Chr09
PMGC_2803	XVII	XVII	-	XVII	scaffold 9	Chr09
PMGC_2515	IV	IV	-	XIV	scaffold 14	no hit
GCPM_3893-1	V	-	-	VII	scaffold 7	scaffold_139&Chr07
ORPM_488	VI	-	-	XVIII	scaffold 18	Chr18
PMGC_2861	XI	-	-	scaffold_170	scaffold 9	Chr09
PMGC_433	XVI	-	XVI	scaffold_125	scaffold 19	scaffold_141&Chr19
scaffold_123_1_361	XVII	-	-	scaffold_123	scaffold 5	Chr17
bu890808	III	-	-	III	scaffold 22	Chr03
ORPM_26	VI	VI	-	VI	no hit	Chr06
scaffold_121_1_246	III	-	-	scaffold_121	scaffold 22	Chr03

L'ordre des marqueurs dans ce cluster de distorsion est en accord avec l'ordre physique sur la séquence de *P. trichocarpa* (**Figure III.5**). Cette zone est monomorphe chez la femelle.

Comparaison avec le génome de P. trichocarpa

Les marqueurs SSR et SNP utilisés nous ont permis d'ancrer les deux cartes parentales sur le génome de *P. trichocarpa* (**Figure III.5**). Une bonne colinéarité entre les cartes génétiques de *P. nigra* et la carte physique du génome de référence *P. trichocarpa* est respectée sauf dans certains cas où la position des marqueurs ne correspond pas aux chromosomes / scaffolds correspondants (**Tableau III.5**). La comparaison des localisations des marqueurs sur les groupes de liaison avec les trois versions de l'assemblage de *P. trichocarpa* montre que les trois versions sont complémentaires. Parmi les discordances avec une au moins des 3 versions d'assemblage répertoriées dans le **tableau III.5**, il existe 7 concordances entre notre carte génétique et la version 1.0, 6 concordances avec la version 3.0 et seulement 3 concordances avec la version v 2.0. Les trois versions d'assemblage étaient concordantes entre elles dans 4 cas. Le marqueur ORPM_269 sur le groupe de liaison XI concorde avec les trois versions du génome. Le PMGC_2550 du groupe de liaison IV, le GCPM_3893-1 du groupe de liaison V et l'ORPM_488 du groupe de liaison VI ne correspondent pas aux chromosomes 1, 7 et 18 des trois versions d'assemblage (**Tableau III.5**).

Comparaison de nos cartes avec les cartes P. nigra de la littérature:

Carte P. nigra (Gaudet et al. 2008)

L'ordre des marqueurs est respecté dans tous les cas sauf pour 1 marqueur sur le groupe de liaison XI. La position du marqueur GCPM_943 est inversée par rapport à PMGC_333. Tout en sachant que Mapmaker classe GCPM_943 comme marqueur accessoire (sa position n'est pas très robuste) donc cette différence est peut être due à une incertitude sur la position de GCPM_943 sur la carte de Gaudet et al. (2008).

Il existe 2 marqueurs dans la carte *P. nigra* (Gaudet et al. 2008) qui sont bi-locus, on trouve le marqueur ORPM_269 sur les groupes de liaison XI et VIII alors qu'il est mono-locus sur le groupe de liaison XI dans nos cartes. De même le marqueur ORPM_16 est localisé sur deux locus appartenant au groupe de liaison XIII dans la carte *P. nigra* (Gaudet et al. 2008), alors qu'il est mono-locus dans nos cartes sur le groupe de liaison XIII (**Tableau III.5**)

Le marqueur PMGC_2550 se trouve sur le groupe de liaison IV dans la carte femelle alors qu'il est sur le groupe de liaison I chez *P. trichocarpa* et sur la carte *P. nigra* (Gaudet at al. 2008)

Tableau III.6 : Liste des 20 premiers termes GO (Gene Ontology) pour lesquels les chromosomes 6 et 18 du peuplier sont significativement enrichis. Un total de 35 410 modèles de gènes possédant au moins un terme GO ont été utilisés. Les termes GO sont classés en trois catégories : M : fonction moléculaire ; B : processus biologique ; C : composant cellulaire (non représenté ici).

Code GO	Terme GO	Catégorie de GO	Occurrence dans :		P-value
			le génome	le set de gènes étudiés	
Enrichissements pour le chromosome 6 (1-21 Mb ; 2027 gènes / 1731 avec GO)					
GO:0050502	cis-zeatin O-beta-D-glucosyltransferase activity	M	23	10	4.84E-08
GO:0050403	trans-zeatin O-beta-D-glucosyltransferase activity	M	23	10	4.84E-08
GO:0047213	anthocyanidin 3-O-glucosyltransferase activity	M	12	6	9.69E-06
GO:0035251	UDP-glucosyltransferase activity	M	265	30	1.84E-05
GO:0016117	carotenoid biosynthetic process	B	214	24	1.39E-04
GO:0016109	tetraterpenoid biosynthetic process	B	214	24	1.39E-04
GO:0034620	cellular response to unfolded protein	B	356	34	1.69E-04
GO:0035967	cellular response to topologically incorrect protein	B	356	34	1.69E-04
GO:0006986	response to unfolded protein	B	356	34	1.69E-04
GO:0016116	carotenoid metabolic process	B	224	24	2.79E-04
GO:0015020	glucuronosyltransferase activity	M	88	13	3.41E-04
GO:0008142	oxysterol binding	M	16	5	7.71E-04
GO:0006984	ER-nucleus signaling pathway	B	378	33	0.00101236
GO:0004161	dimethylallyltranstransferase activity	M	5	3	0.00108255
GO:0016871	cycloartenol synthase activity	M	5	3	0.00108255
GO:0006723	cuticle hydrocarbon biosynthetic process	B	5	3	0.00108255
GO:0016840	carbon-nitrogen lyase activity	M	26	6	0.00133785
GO:0042177	negative regulation of protein catabolic process	B	18	5	0.001393
GO:0004177	aminopeptidase activity	M	36	7	0.00157791
GO:0080167	response to karrikin	B	298	27	0.00163407
Enrichissements pour le chromosome 18 (1370 gènes / 1188 avec GO)					
GO:0015276	ligand-gated ion channel activity	M	104	20	2.45537E-10
GO:0022834	ligand-gated channel activity	M	104	20	2.45537E-10
GO:0005234	extracellular-glutamate-gated ion channel activity	M	81	17	1.30939E-09
GO:0005231	excitatory extracellular ligand-gated ion channel activity	M	81	17	1.30939E-09
GO:0004970	ionotropic glutamate receptor activity	M	81	17	1.30939E-09
GO:0035235	ionotropic glutamate receptor signaling pathway	B	81	17	1.30939E-09
GO:0007215	glutamate receptor signaling pathway	B	81	17	1.30939E-09
GO:0008066	glutamate receptor activity	M	84	17	2.36906E-09
GO:0006874	cellular calcium ion homeostasis	B	100	17	3.72291E-08
GO:0055074	calcium ion homeostasis	B	104	17	6.77995E-08
GO:0072503	cellular divalent inorganic cation homeostasis	B	140	19	2.47995E-07
GO:0009244	lipopolysaccharide core region biosynthetic process	B	7	5	8.36788E-07
GO:0046401	lipopolysaccharide core region metabolic process	B	7	5	8.36788E-07
GO:0009702	L-arabinokinase activity	M	5	4	6.13527E-06
GO:0006875	cellular metal ion homeostasis	B	146	17	8.75721E-06
GO:0016595	glutamate binding	M	89	17	3.62861E-05
GO:0009103	lipopolysaccharide biosynthetic process	B	13	5	4.33171E-05
GO:0004335	galactokinase activity	B	15	5	9.55607E-05
GO:0006995	cellular response to nitrogen starvation	B	183	17	0.000156846
GO:0003860	3-hydroxyisobutyryl-CoA hydrolase activity	M	10	4	0.000225114
GO:0005267	potassium channel activity	M	50	8	0.000240584

Ce marqueur qui est supposé être un marqueur pont ne se lie dans la carte maternelle qu'avec ORPM_425, un marqueur très distordu (Pχ2 < 0,0001) (**Tableau III.5**).

Carte P. nigra (Cervera et al. 2001):

L'ordre des marqueurs est respecté dans tous les cas. Deux divergences existent quant à l'appartenance aux groupes de liaison, le marqueur WPMS_6 est localisé sur le groupe de liaison XII sur la carte *P. nigra* (Cervera et al. 2001) alors qu'il est sur le groupe de liaison XVI sur nos cartes et sur celle de Gaudet al. 2008. De même, le marqueur PMGC_451 se trouve sur le groupe de liaison C dans la carte *P. nigra* (Cervera et al. 2001). Ce marqueur est localisé sur le groupe de liaison XVII dans les cartes *P. nigra* (Gaudet et al. 2008 ; Elmalki et al. soumis) (**Tableau III.5**).

3. Analyse par enrichissement « Gene Ontology »

Il m'a semblé intéressant d'analyser plus précisément deux régions du génome particulières : une grande partie du groupe de liaison VI qui est distordu et le groupe de liaison XVIII, chez qui plusieurs marqueurs testés se sont tous révélés monomorphes. Un enrichissement des termes GO a donc été entrepris dans ces 2 régions (**Tableau III.6**). Le groupe de liaison VI est significativement enrichi pour 56 termes GO (*P-value* < 0.001 ; données non montrées). Les gènes associés aux termes GO enrichis ont été analysés par bibliographie. Seuls ceux associés au terme GO « cycloartenol synthase activity » (GO:0016871) ont retenu notre attention du fait de l'existence de mutants létaux chez *Arabidopsis thaliana*. L'analyse du groupe de liaison XVIII montre un total de 78 termes GO significativement enrichis sans qu'il soit possible de faire ressortir des termes particuliers liés à la distorsion de ségrégation observée sur ce groupe de liaison.

D. Intérêt de ces nouvelles cartes génétiques (Discussion)

1. Ancrage de la carte génétique du *P. nigra* sur la carte physique de *P. trichocarpa* et intérêt pour l'assemblage *de novo* de *P. nigra*

L'ancrage de la carte génétique *P. nigra* sur le génome de *P. trichocarpa* a permis de révéler une proportion élevée de synthénie et de colinéarité. Malgré leur spéciation, les deux espèces ont conservé une colinéarité entre leurs génomes. Il existe néanmoins des différences entre certains marqueurs de notre carte et leur localisation sur le génome de *P. trichocarpa*. Cela pourrait être dû à l'existence d'un grand nombre de zones dupliquées du génome. Le génome du peuplier a été dupliqué il y a 65 millions d'années et on estime à 92% la taille des régions dupliquées sur le génome (Tuskan et al. 2006). Hormis, le marqueur PMGC_2550 qui est lié à

marqueur très distordu et se retrouve sur le groupe de liaison IV, les deux autres marqueurs qui sont en discordance avec les trois versions d'assemblage correspondent à des groupes de liaison largement dupliqués chez le peuplier : le marqueur ORPM_488 appartenant aux groupes de liaison VI et XVIII. Le marqueur GCPM 3893-1 appartenant aux groupes de liaison V et VII (Tuskan et al. 2006). Une autre raison de divergence est due aux différentes versions d'assemblage du génome. Malgré le fait que la version 3.0 du *P. trichocarpa* a intégré 81 Mb de séquences supplémentaires et s'est basée pour l'assemblage sur une carte génétique de 19 groupes de liaison, elle n'est toujours pas complète. Néanmoins, les versions 1, 2 et 3 demeurent complémentaires comme illustré par leur comparaison avec nos cartes génétiques (**Tableau III.5**).

Outre les STS (Sequence Tagged Sites) qui sont des séquences uniques dans le génome, les cartes génétiques permettent de disposer de points de repère, ordonnés le long des chromosomes par la mesure de leur liaison deux à deux. En 1996, le Généthon a publié une carte génétique de référence du génome humain, ordonnant 5264 microsatellites, qui est encore très utile aujourd'hui (Dib et al. 1996). Elle a notamment permis de cartographier de nombreux gènes associés à des maladies génétiques et a servi pour l'assemblage de la séquence du génome humain. Chez *P. trichocarpa*, 356 SSR ont été utilisés pour assigner 155 scaffolds y compris les 19 chromosomes. La majorité des SSR cartographiés (91%) sont colinéaires avec la séquence physique. En combinant la carte physique et carte génétique, 385Mb des 410Mb ont été assemblés dans des chromosomes (Tuskan et al. 2006).

Après ancrage sur la séquence de *P. trichocarpa*, notre étude avait montré un scaffold mal assigné dans la version 2.0. Il s'agissait du scaffold 22 (606 585 pb) qui est ancré au groupe de liaison III sur notre carte. La version 3.0 du génome a permis de confirmer cela puisque désormais ces marqueurs appartiennent au scaffold 3. Cet exemple nous illustre la plus grande utilité des cartes génétiques pour l'assemblage de contigs et de super-contigs avant l'établissement de cartes physiques. Les marqueurs moléculaires des cartes génétiques restent un outil précieux pour valider et ancrer les cartes physiques le long des chromosomes. Le séquençage *de novo* des génomes *P. nigra* pourrait se baser sur l'ancrage sur la séquence de *P. trichocarpa*, et sur les deux cartes génétiques *P. nigra* existantes (Gaudet et al. 2008) et la carte consensus de notre présent travail.

2. Génétique comparée peuplier/saule

L'utilisation des marqueurs SSR multi-alléliques et le taux de transférabilité élevé intra-espèce nous a permis de comparer nos cartes avec les cartes *P. nigra* (Gaudet et al. 2008 et Cervera et al. 2001). Cette comparaison nous a confirmé la bonne qualité de nos cartes.

Le succès de transférabilité des marqueurs SSR chute en inter-espèce et chute drastiquement quand on les teste sur le saule. La comparaison des cartes peuplier et saule ne peut se faire alors avec ce genre de marqueurs. Pour ancrer les cartes génétiques du saule sur le génome du peuplier, des marqueurs SNP appartenant à des modèles de gènes du peuplier ont été développés. Pour avoir des SNP pouvant s'ancrer dans le génome de *P. trichocarpa*, Hanley et al. (2007) ont ciblé les EST cartographiés sur le génome du peuplier et ont développé les amorces à partir de ces fragments (Hanley et al. 2007) alors que Berlin et al. (2010) ont privilégié le ciblage des gènes candidats pour les caractères d'intérêt comme la résistance à la rouille. Les deux travaux ont utilisé aussi quelques marqueurs SSR développés sur le peuplier. Grâce à tous ces marqueurs, ces deux cartes ont mis en évidence une macro-synthénie élevée entre les deux espèces et un ordre de gènes respecté à travers le génome. Le taux de mutations silencieuses dans les gènes orthologues est trois fois moins prononcé que dans les gènes paralogues suggérant une forte conservation de fonctions de gènes entre les deux espèces (Berlin et al. 2010).

3. Distorsion

Les cartes génétiques peuvent révéler des évènements de sélection méiotiques (durant la gamétogénèse) ou post-méiotiques (durant la pollinisation, la formation des graines, la germination des graines et lors de la croissance de la plante). Lors de la construction de cartes génétiques, des distorsions de ségrégation plus ou moins importantes peuvent se révéler. Lorsque leur nombre est supérieur à celui attendu par le seul fait du hasard et lorsqu'elles sont regroupées sur un même segment chromosomique, on a pu mettre en évidence des liaisons étroites entre locus montrant des ségrégations distordues et des locus d'incompatibilité (combinaisons alléliques incompatibles), des gènes de létalité à l'état récessif (Bradshaw and Stettler 1994) et des réarrangements chromosomiques (Kianian and Quiros 1992).

Dans notre cas, le groupe de liaison VI, qui est fortement distordu chez le parent mâle, présente un enrichissement du terme GO « cycloartenol synthase activity ». Des mutants létaux (CAS1) ont été identifiés chez *A. thaliana*. Une mutation (cas1-2) au niveau de ce gène induit un arrêt de croissance et une mort de l'individu par nécrose (Babiychuk et al. 2008). Chez le mâle *P. alba*, cette partie du groupe de liaison VI n'a pas pu être obtenue par manque de polymorphisme (Paolucci et al. 2010). Par ailleurs, les groupes de liaison VI et XVIII sont dupliqués. Les

marqueurs présents sur le groupe de liaison XVIII sont particulièrement monomorphes chez les deux parents *P. nigra* alors qu'un cluster de marqueurs distordus est détecté chez la femelle *P. alba* (Paolucci et al. 2010).

4. Conclusion : intérêt des cartes pour la cartographie de QTL

Les cartes jalons construites sont utilisées dans les chapitres suivants pour la détection de QTL contrôlant la résistance partielle à *Mlp* et les composantes de la qualité de la biomasse. Le calcul de la couverture selon Hulbert et al.(1988) donne des valeurs assez faibles de couverture pour les cartes jalons femelle (35%) et mâle (44%). Toutefois, sur la base de l'ancrage de ces cartes sur la séquence du génome (**Figure III.5**), ces cartes couvrent une grande partie du génome séquencé. On peut estimer que ces cartes sont satisfaisantes pour la détection de QTL.

- Cartographie de QTL associés à la résistance partielle à la rouille foliaire

Chapitre IV : Cartographie de QTL associés à la résistance partielle à la rouille foliaire

La rouille du peuplier, principalement causée par le champignon pathogène *Melampsora larici-populina*, est aujourd'hui la maladie la plus préoccupante pour la populiculture européenne. Plusieurs contournements de résistance aux conséquences économiques majeures montrent les limites des stratégies de sélection développées en Europe depuis 30 ans, celles-ci étant exclusivement orientées vers la valorisation de résistances complètes apportées par l'espèce parente *P. deltoides* (Pinon and Frey 2005). Des efforts de recherche conséquents ont porté récemment sur une meilleure connaissance des bases génétiques des résistances partielles du peuplier face à la variabilité connue de *Melampsora larici-populina* (Dowkiw and Bastien 2004, 2007; Jorge et al. 2005). En n'interdisant pas la maladie, ces résistances partielles limitent en théorie les possibilités d'évolution du champignon. Connaître leur déterminisme génétique en croisement intra et interspécifique et leurs liens avec les résistances complètes est indispensable afin de cumuler dans un même individu tous les mécanismes naturels de défense disponibles.

Le développement de techniques d'inoculation en conditions de laboratoire par une souche unique de *Melampsora* permet de contrôler la présence de résistances complètes à une ou plusieurs races du pathogène et de décomposer la résistance partielle en trois paramètres : latence infectieuse, niveau d'infection apprécié par le nombre de sores à 14 jours et la sporulation appréciée par la taille relative des sores (Dowkiw et al. 2003). C'est pour l'intensité de sporulation que le niveau de variabilité génétique de la plante hôte est le plus élevé et que la variabilité de réponse est la plus faible lorsque la variabilité de l'agent pathogène est prise en compte. C'est d'ailleurs cette dernière composante évaluée sur un seul cycle d'infection qui est la mieux corrélée avec le niveau d'attaques en conditions naturelles d'infection (Legionnet et al. 1999; Dowkiw et al. 2003). Cette résistance de nature quantitative est principalement héritée des parents *P. trichocarpa* et *P. nigra.* Elle peut être dans certains cas cumulée efficacement à un effet résiduel d'une résistance complète contournée par l'agent pathogène (Dowkiw and Bastien 2007).

Les résultats acquis sur le déterminisme génétique et la souche-spécificité des résistances complètes et partielles aux rouilles foliaires conduisent les sélectionneurs du peuplier à s'interroger sur les réelles possibilités de développement de résistance durable au sein d'hybrides interspécifiques (*P. deltoides x P. trichocarpa*). Il convient aujourd'hui d'explorer un ensemble de solutions permettant de limiter les pressions de sélection exercées sur les populations de l'agent

pathogène *Melampsora larici-populina* et ainsi réduire les possibilités d'adaptation de ce dernier. Un intérêt plus grand est porté à valorisation de sources de résistance partielle présente dans l'espèce *P. nigra* du fait d'une co-évolution de longue date de cette espèce avec *M. larici-populina*, pathogène à aire de distribution européenne.

Toutes les études de détection de régions génomiques impliquées dans le contrôle de la résistance partielle à *Mlp* en conditions d'infection naturelle ou artificielle en laboratoire ont concerné des pédigrées de cartographie interspécifiques n'impliquant pas l'espèce *P. nigra* (cf chapitre III). C'est pourquoi, nous avons voulu valoriser la famille de cartographie disponible pour explorer la complexité de l'architecture génétique de la résistance partielle à *Mlp* dans cette espèce. Pour la détection de QTL réalisée dans ce travail de thèse, j'ai été conduit à analyser la variabilité génétique observée pour trois composantes épidémiologiques après inoculation avec deux souches différentes de *Mlp*, ces tests ayant été réalisés avant mon arrivée au laboratoire. Les résultats de cette analyse de données phénotypiques et de la détection de QTL associés aux différentes composantes de la résistance partielle ont fait l'objet d'un projet de publication soumis à la revue PlosOne présenté ci-après.

En résumé, une importante variabilité génétique a été observée pour les trois composantes épidémiologiques de la résistance partielle dans les deux tests d'inoculation. Une forte interaction génotype x souche a été identifiée chez le parent mâle du pédigrée (BDG) ainsi que chez la plupart des descendants. Cette interaction s'explique non seulement par une plus grande agressivité de la souche 98AG69 permettant l'expression d'une ségrégation de nature bimodale de la taille moyenne des sores mais aussi par un changement de classement des génotypes pour les trois composantes épidémiologiques étudiées. Un total de 11 QTL répartis sur 8 chromosomes et expliquant entre 2% et 60% de la variation phénotypique observée après inoculation avec les deux souches, ont été mis en évidence. Tous ces QTL sont souche ou trait dépendants sauf le QTL à effet fort détecté à l'extrémité du groupe de liaison°I. Ce QTL majeur hérité du parent mâle explique plus de 60% de la variation de la taille des sores après inoculation avec la souche 98AG69 et montre un effet modéré après l'inoculation avec la souche 93JE3. Ce QTL co-localise avec un QTL majeur de résistance à la rouille foliaire à *Melampsora larici-epitea* identifié chez le saule (*Salix schwerinii*) (Samils et al. 2011). L'exploration génomique de l'intervalle de confiance autour de ce QTL a révélé un cluster de 25 gènes NB-LRR (Nucleotide Binding Leucine-Rich Repeat). La complexité du contrôle génétique de la résistance partielle à *Mlp* révélée chez *P. nigra* est discutée pour une meilleure valorisation de cette résistance en sélection.

QTL mapping in *Populus nigra* reveals high strain specificity of partial resistance against the coadapted pathogen *Melampsora larici-populina*.

Redouane El Malki[1], Véronique Jorge[1], Vincent Segura[1], Rémi Bounon[2], Aurélie Bérard[2], Patricia Faivre-Rampant[3], Vanina Guérin[1], Patrick Poursat[4], Arnaud Dowkiw[1] and Catherine Bastien[1*]

[1] INRA, UR0588, Amélioration Génétique et Physiologie Forestières, F-45075 Orléans, France.

[2] INRA, US1279, Étude du Polymorphisme des Génomes Végétaux, CEA Institut de Génomique, F- 91057 Évry, France.

[3] INRA, UMR1165, Unité de Recherches en Génomique Végétale, F- 91057 Évry, France.

[4] INRA, UE0995, Génétique et Biomasse Forestières Orléans, F-45075 Orléans, France.

Abstract

Fast-growing poplar clones have a high potential of ligno-cellulosic biomass production in short-rotation coppices (SRC). Nevertheless, with high plant density, the SRC system represents a wet environment favoring epidemic diseases such as leaf rust caused by the Eurasian fungus *Melampsora larici populina (Mlp)*. Natural populations of *Populus nigra (P. nigra)* act as the main reservoir of genetic variability for durable leaf rust partial resistance. Deployment of inter-specific poplar clones (*P. deltoides x P.nigra*) is today the breeding strategy adopted to achieve durable resistance in plantations. The QTL detection performed in this study brings out an initial insight on the genetic inheritance of leaf rust partial resistance in *P. nigra*. Partial resistance to *Mlp* was investigated in 324 full-sib clones from a *P. nigra* family after artificial inoculation in leaf-disk bioassays. Partial resistance against two different strains was evaluated with three epidemiological components: latent period, uredinia number and uredinia size. The genotyping with SSR, SNP and AFLP markers of the entire progeny and the two parents allowed construction of two parental maps and one consensus map anchored on the reference *P. trichocarpa* genome. Synteny and colinearity with the reference *P. trichocarpa* genome and published *P. nigra* genetic maps were respected.

High genetic variation was observed for partial resistance to the two *Mlp* strains. Significant strain specificity of resistance response was detected for one parent and most of their progeny. A total of eleven *QTL* distributed over eight chromosomes explained between 2% and 60 % of the phenotypic variation observed after inoculation with the two strains. Most of these QTL were trait- or strain-specific except one. This major QTL inherited from the most resistant parent and located at the end of LG I explained more than 60% of uredinia size variation after inoculation with one *Mlp* strain and showed a moderate effect after inoculation with the other strain. The genomic mining of the QTL region which is co-located with a major QTL associated to Salix resistance to *Melampsora larici-epitea* revealed a cluster of 25 NB-LRR genes. The relative complex genetic control of partial resistance to *Mlp* revealed in *P. nigra* is discussed in relation to the new perspectives offered in breeding for durable resistance.

Key words: QTL mapping, poplar, leaf rust resistance, *Melampsora larici populina*

A. Introduction

Poplars are among the fastest growing trees at temperate latitudes and are of considerable commercial importance (Sannigrahi et al. 2010). Most cultivars grown in Europe are interspecific hybrids between the two North American species *Populus deltoides* and *P. trichocarpa* or between *P. deltoides* and the European black poplar *P. nigra*. Poplar timber has many uses including paneling, packaging and pulp. Being a fast-growing tree, poplar can be harvested early, thus having a potential as a short-rotation coppice (SRC) crop grown for biomass energy (Berthelot and Gavaland 2007). With plant densities ranging from 2,500 to more than 10,000 plants/ha, SRC represent a wet environment potentially favoring fungal diseases. One of the most important fungal pathogens threatening poplar stands in northern Europe is *Melampsora larici populina Kleb.* leaf rust (abbreviated here as *Mlp*). *Mlp* is an obligate macrocyclic basidiomycete which undergoes sexual reproduction on larch (*Larix* sp.) and replicates asexually on poplars until leaf fall. Although the pathogen originates from Eurasia, all three poplar species mentioned earlier happen to be susceptible to the disease, *P. deltoides* is a source of both partial and complete resistances whereas the two other species, including the co-evolved species *P. nigra*, seem to exhibit only partial resistance (Pinon 1992; Pinon and Frey 2005; Legionnet et al. 1999). However, most of the completely rust-resistant poplar cultivars released so far by European poplar breeders and that carried complete resistances inherited from *P. deltoides* became susceptible to *Mlp* in less than ten years after commercial deployment (Pinon and Frey 2005). Breeding poplars with durable resistance to *Mlp* is a real challenge, especially in a context where *Mlp* populations exhibit high genetic diversity in black poplar-larch sympatric areas (Xhaard et al. 2012) and where selective pressures from the domesticated host are both rapid and strong (Xhaard et al. 2011). Most European breeding programs are now focusing their efforts in exploiting the genetic variation for partial resistance, especially in the co-evolved species *P. nigra* (Giordano 1973; Gergacz 1978; Krzan 1981; Schoeneman 1986; Slycken et al. 2002).

The genetic architecture of *Mlp* resistance in poplar is well described for the non co-evolved species *P. deltoides* and *P. trichocarpa*. (Jorge et al. 2005; Bresson et al. 2011; Cervera et al. 1996; Zhang et al. 2001). Both complete resistance (*P. deltoides*) and partial resistance (*P. deltoides* and *P. trichocarpa*) factors have been mapped in segregating progenies with increasing precision using bulked-segregant analysis, QTL detection and, more recently, fine mapping. Interestingly, all findings contributed to blur the lines between complete and partial resistances : (i) two complete resistance factor from *P. deltoides*, *Mer* (Zhang et al. 2001) and R_1

(Bresson et al. 2011) as well as a major partial resistance factor from *P. trichocarpa*, R_{US} (Bresson et al. 2011), were finely mapped into NB-LRR rich regions of chromosome XIX, (ii) R_{US} exhibited extreme strain-specificity resembling that of complete resistance (Dowkiw et al. 2010), (iii) defeated complete resistances from *P. deltoides* seemed to act as partial resistance QTL against compatible *Mlp* strains (Jorge et al. 2005; Dowkiw and Bastien 2007), and (iv) most minor partial resistance QTL identified were specific not only of a subset of strains but of a given resistance component (latent period, uredinia size or uredinia number) for a given strain (Jorge et al. 2005).

By contrast, very little is known about the genetic architecture of partial resistance to *Mlp* in the co-evolved species *P. nigra*. Two early studies (Pinon and Du Cros 1976; Teissier du Cros 1977) where *P. nigra* female trees from European and French populations, respectively, plus their respective open-pollinated progenies were evaluated for field resistance to *Mlp* revealed significant parent-offspring and inter-annual correlations, thus suggesting the existence of heritable genetic factors in the species. By contrast, using a factorial mating design where a set of male and female *P. nigra* genitors and their progenies were evaluated for field resistance, a later study (Pichot and Teissier du Cros 1988) concluded that most of the variance lies within progenies, and thus that the resistance level of an individual cannot be inferred from its parents' performances. The fact that the parents involved in this work had been phenotypically selected for vigour and form may explain part of the discrepancy between both studies. The results from Pinon and Teissier du Cros (1976) suggested a significant but unpredictable geographical pattern for rust resistance within *P. nigra*. Individuals from the French Drôme and Rhône rivers were found to be more resistant than those from the French Alps while Hungary accessions were the most susceptible. A later study involving field tests and inoculated leaf-disc bioassays on 283 genotypes from various French populations also revealed significant but trait-dependant and still unpredictable geographical patterns (Legionnet et al. 1999). However, the variation attributable to the geographical structure was small relative to individual variation. The fact that most resistance components measured in this study were found to be uncorrelated raised hopes about the possibility of combining different genetic factors from *P. nigra* into a single variety, and thus of achieving more durable resistance.

Although genetic maps of *P. nigra* have been developed elsewhere, they were generated either from an interspecific *P. deltoides* x *P. nigra* mapping progeny (Cervera et al. 2001) or from an intraspecific mapping progeny (Gaudet et al. 2008) whose parents were found to be highly susceptible to *Mlp* in our laboratory (unpublished data). In order to perform this unprecedented

QTL detection for *Mlp* resistance in *P. nigra*, we generated our own *P. nigra* maps (one for each parent plus a consensus map) using segregating progeny from two heterozygous parents with contrasted resistance levels. This work benefited from the populus genome sequence (Tuskan et al. 2006, http://www.phytozome.org/) for the design of SNP markers in known genes, for anchoring the maps and for localizing the QTL. Based on the hypothesis that partial resistance to *Mlp* is under complex genetic determinism in *P. nigra*, both strain-specificity and epistasis were considered. Evaluating partial resistance in laboratory bioassays allowed decomposing this integrative trait into more simple – and possibly genetically independent - epidemiological components.

B. Material and Methods

1. Plant material

A pseudo-testcross F_1 segregating population was generated from a cross between two contrasted *P. nigra* parents. The '71072-501' female parent was sampled in 1972 in a wild population from South-Eastern France (45°35' N, 05°36' E). The male parent, 'Blanc de Garonne' (abbreviated here as 'BDG') is a cultivar originating from a phenotypic selection made before 1950 in wild South-Western French populations. 'BDG' was classified as "rather susceptible" under natural infection by a mixed population of *Mlp* and *Melampsora allii-populina* Kleb in an early study (Lemoine and Pinon 1978). The field and laboratory observations that we have accumulated on our collection of more than 1,000 *P. nigra* clones suggest that 'BDG' is in fact much more resistant than '71072-501' (unpublished data).

A total of 324 full-sib cloned genotypes were used in this study: 19 individuals originating from an initial controlled cross made in 1994 as part of a 5x4 factorial mating design, and 315 individuals originating from a second controlled cross made in 2002. Although these 324 genotypes served for genetic map construction, resistance assessment was restricted to the 308 genotypes that produced the amount of leaf material necessary for resistance evaluation. This progeny was maintained as mother stocks in our nursery (47°46'N, 1°52'E, 110m). All F_1 clones plus the two parents were grown from cuttings in 3-L pots containing 20% sand, 40% peat and 40% ground bark mixture under glasshouse rust-free conditions.

2. Rust resistance assessments

Inoculation tests were conducted on 3-cm-diameter excised leaf discs sampled from the fifth to the tenth unrolled leaf below the apex. The leaf discs were floated upside down on distilled water in six-well culture plates and inoculated with single strain isolates. Two contrasted *Mlp* strains

were used: 98AG69 and 93JE3. Strain 98AG69 was isolated from *P. deltoides x P. trichocarpa* cv. 'Beaupré' in 1998 and belongs to pathotype 1-3-4-5-7, thus harbors five of the eight virulences that have been defined so far. It is a reference strain used at INRA. Strain 93JE3 was isolated from *P. deltoides x P. nigra* cv 'Blanc du Poitou' in 1993 and belongs to pathotype 2-3-4, which is more simple and thus more similar to what is usually found in natural *P. nigra* stands. Among other known differences, 98AG69 develops small uredinia in the presence of R_{US}, a major partial resistance factor inherited from *P. trichocarpa* whereas 93JE3 was the first identified R_{US}-defeating strain (Dowkiw et al. 2010).

Inoculations were performed according to Dowkiw et al. (2003). For inoculation with one given strain, the experimental design was a five-randomized-complete-block design. Each of the 308 F_1 genotypes was represented by one disk per block and each of the two parents by 3 disks per block. Mono-strain urediniospore suspensions were sprayed over the discs with a hand atomizer. The actual number of uredinia deposited per leaf disk was estimated by scattering Petri dishes containing solid agar (20 g/liter) in the experiments before inoculation and by counting the number of deposited urediniospores using a microscope. The estimated values were 236 (SD = 83.55) for 93JE3 and 187 (SD = 49.26) for 98AG69. Average germination rates measured 48h after inoculation on solid agar were respectively 87% for 93JE3 and 81% for 98AG69.

The disks were kept for 14 days under controlled conditions (17°C, 16h-photoperiod). Three epidemiological components were measured on each disk: latent period (LP) measured twice a day as the time after inoculation when the first sporulating uredinia appeared, number of uredinia (UN) at 13 days after inoculation, and uredinia size (US) at 14 days after inoculation. US was scored on a 1 (small) to 5 (large) ranking scale.

3. Genotyping

Genomic DNA was extracted from fresh leaf tissue using the DNeasy 96 Plant Kit (Qiagen). Three types of genetic markers were used. Priority was given to SSR and SNP markers to ensure bridges across the two generated parental maps but also with the reference *Populus* physical map (*P. trichocarpa* 'Nisqually-1', (Tuskan et al. 2006)) and with other published genetic maps on different *Populus* species (Cervera et al. 2001; Jorge et al. 2005; Street et al. 2006; Gaudet et al. 2008; Paolucci et al. 2010). AFLP markers were then used to generate rapidly large number of mappable markers.

SNP markers

SNP markers were identified through analyzing sequences of 665 gene fragments in the two parents. Although these fragments were candidate genes for resistance to biotic and abiotic stresses, growth or phenology, they were most importantly spread over the whole genome. Sequences were analyzed using the software Genalys (Takahashi et al. 2003). Uniqueness of SNP targets was checked by Blastn against the reference *P. trichocarpa* genome sequence. SNP markers were then selected based on an absence of polymorphism in neighboring regions (15 bp on each side). Due to the double pseudo-testcross mapping strategy implemented here, SNP markers that were polymorphic for both parents were not fully informative and not numerous enough, and thus they were not retained. We however implemented a strategy to make bridges across the two parental maps. It consisted in selecting in the same gene fragment, some SNP in pairs so that each one was polymorphic for a distinct parent. Given the size of the analyzed fragments (<600 pb), less than one recombinant genotype was expected within the studied progeny. Consequently, each pair of SNP could be considered as a single multi-allelic locus segregating in four classes within the pedigree . A total of 13 SNP pairs were found that met our requirements. An additional seventy "unpaired" SNP markers were found. SNP genotyping was performed using two techniques: PAMSA (PCR amplification of multiple specific alleles (Gaudet et al. 2007) for the SNP belonging to the first three scaffolds, and SNPLex^{TM} (Applied Biosystems) for 96 SNP selected on the other scaffolds. The original PAMSA protocol has been modified as following: amplification products were labeled with fluorescence and separated on an ABI 3100 16-capillary DNA sequencer (PE Applied Biosystems, Foster City, Calif.). Extending the primers by only three nucleotides was thus sufficient to distinguish SNP alleles (*vs* 5bp and 15bp long tails in the original protocol). For the SNPLex^{TM} technology, more details are available on the Web (http://www3.appliedbiosystems.com/cms/groups/ mcb_support/documents/generaldocuments/cms_042019.pdf).

SSR markers

SSR primers were chosen from two different sources: (1) IPGC (http://www.ornl.gov/sci/ipgc/ssr_resource.htm), for markers named with prefix "PMGC", "ORPM" (Tuskan et al. 2004) or "WPMS" (Schoot et al. 2000); and (2) data mining from the public *Populus* sp. EST databases for markers named with prefix "bi" or "bu" (Monclus et al. 2012). SSR polymorphism was checked on a subset of 6 progenies plus the two parents. Polymorphic SSR were then selected for genotyping the entire progeny. Two techniques were successively used for fluorescent labeling: fluorochrome- labeled forward primer or M13 tailing which reduces labeling costs (Schuelke 2000). The first technique was implemented as

described in Jorge et al. (2005) and the M13-tailing technique as described in Monclus et al. (2012). SSR allele lengths were recorded with the GeneMapper® software (Applied Biosystems).

AFLP markers

AFLP genotyping was based on the EcoR1 (E) - *Mse*I (M) restriction enzyme pair. The first PCR (pre-amplification) was conducted with one selective nucleotide (E+A and M+C). It was followed by a selective amplification with three selective nucleotides (E+ANN and M+CNN) where the E+ANN primer was labeled with 6FAM or HEX. The PCR products were analyzed on an ABI 3100 16-capillary DNA sequencer (PE Applied Biosystems, Foster City, Calif.). Polymorphism was checked on a subset of 6 progenies plus the two parents. Primer combinations maximizing the number of polymorphic AFLP markers were then selected for genotyping the entire progeny. AFLP markers were named "E*ann*M*cnn*" followed by the size of the fragment in bp.

4. Genetic map construction

Markers significantly deviating from a [1:1] Mendelian segregation ratio (*i.e.* $P_{\chi^2} \leq 0.01$) were considered as distorted. To allow identifying possible regions of distortion, all markers were used to build the two parental maps plus the consensus map. Distorted markers were however excluded from the analysis when generating the corresponding three framework maps.

Parental maps

Computations were made with Mapmaker version 3.0 (Lander et al. 2009) using the two-way pseudo-testcross strategy which led to the construction of two parental maps (Grattapaglia and Sederoff 1994). To detect linkages in repulsion phase, the data set was inverted and added to the original data. As a start, markers were grouped by two-point analysis using a logarithm of odds (LOD) score of 4.0 and a maximum recombination fraction of 0.3. The most likely order of markers within a linkage group was then determined by multi-point analysis as follows. For linkage groups with more than five markers, the "three point" command was used to pre-compute the likelihood of all three-point crosses of each group. Then, the "order" command was used to select a subset of markers ordered at a minimum LOD of 4.0 with a log-likelihood threshold of 3.0. Additional markers were added by the "try" command with a log-likelihood threshold of 2.0. The order of the marker subset was controlled with the "ripple" command. New markers were added only if the new order obtained was confirmed with this command. For the linkage groups with less than five markers, the "compare" command was used. The marker order

of these groups was equally supported by a log-likelihood of 2.0. Distances between marker loci were calculated from recombination fraction using Kosambi's mapping function.

Consensus map

Computations were made with JoinMap version 4 (Van Ooijen 2006) under the *cross-pollination* model (CP). Grouping trees were constructed based on "independence test LOD score" and the software's default settings. For all the linkage groups, a LOD grouping threshold of 6.0 was applied. However, when physical position of markers was indicative of independence, this LOD threshold was increased to 16.0. Map distances were calculated using the Kosambi mapping function and the regression mapping algorithm (Stam 1993) with the following parameters: recombination threshold of 0.40, ripple value of 1, jump threshold of 5.0.

Framework maps

Computations were made with JoinMap version 4 (Van Ooijen 2006) to construct two parental plus one consensus framework map. Markers showing distorted segregation ratios were filtered out from the dataset. Grouping trees were constructed based on "independence test LOD score" and the software's default settings. For all the linkage groups, a LOD grouping threshold of 4.0 was applied. However, when physical position of markers was indicative of independence, this LOD threshold was increased to 6.0. Map distances were calculated as above for the consensus map, except that only first or second rounds of placing markers were trusted.

All generated maps were drawn with MapChart 2.1 (Voorrips 2002).

5. Resistance data analysis

Rust resistance data were analyzed using the R software (R Development Core Team) and the ASReml-R package (Butler et al. 2009).

We first analysed epidemiologic data corresponding to each inoculated strain separately. Prior to analysis of variance, the individual data were transformed with the BoxCox procedure to ensure homocedasticity and normality of the residual distribution from the following linear model:

$$Y_{ijk} = \mu + B_i + G_j + \varepsilon_{ijk}$$

where μ is the grand mean, B is the block effect (fixed), *G* is the genotype effect (random), and ε is the residual random error. Accordingly, LP and UN were transformed into LP^{-2} and $UN^{1/2}$. Block effects were significant (P < 5%) for all variables. Genotypic means and variance

components estimates were therefore computed from block-adjusted individual values. Restricted maximum likelihood (REML) estimates of the genotypic (σ^2_G) and residual (σ^2_ε) variances allowed computation of broad-sense heritabilities at individual level as follows: $H^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_\varepsilon)$. Standard errors of H^2 were derived from the covariance matrix of the estimated variance components using the delta method (Lynch and Walsh 1998).

Correlations between LP, UN and US within and between *Mlp* strains were estimated using Spearman's rank correlation coefficients computed on adjusted genotypic means.

To investigate potential genotype-by-strain interactions for a given trait, the block-adjusted data for both *Mlp* strains were combined and analyzed using the following linear mixed model:

$$Y_{ijk} = \mu + S_i + G_j + (GS)_{ij} + \varepsilon_{ijk}$$

Where μ is the grand mean, S is the strain effect (fixed), G is the genotype effect (random), (GS) is the random genotype-by-strain interaction, and ε is the residual random error. Because variances for a given trait can significantly differ between strains, we allowed the model to fit distinct residual variances for each strain. The relevance of this model was assessed through comparison with a "classic" interaction model assuming a unique residual variance using a likelihood ratio test. Variance components REML estimates were computed from the most relevant model. Associated standard-errors estimates were derived from the covariance matrix of the estimated variance components using the delta method (Lynch and Walsh 1998).

6. QTL detection

QTL detection was based on the three framework maps (two parental plus one consensus). Thus, QTL detected on the female and male maps result from respectively female and male additive effects. The consensus map being based on markers that are polymorphic for both parents, up to four allele classes can be defined thus allowing the detection of dominant male-by-female interaction effects in addition to female and male additive effects.

Putative QTL were identified by Interval Mapping (IM) as implemented in MapQTL® 6.0 (Van Ooijen 2009). A 1,000-permutation test was applied to each data set (Churchill and Doerge 1994) to determine LOD thresholds. The 95th percentile of the resulting LOD distribution was used as a genome-wide significance threshold. The chromosomal location with the highest LOD score was considered the most likely QTL position. The significant markers were selected as initial cofactors for the automatic cofactor selection procedure aimed at identifying the most informative markers. The Multiple-QTL model Mapping (MQM) method was subsequently used

Table IV.1 : Summary of number of AFLP, SSR and SNP markers tested and genotyped in the F1 *P. nigra* mapping pedigree.

# Markers	AFLP	SNP	SSR
Tested	101[a]	111	538
Polymorphic	41	105	179
Success rate %	41	94	67
Inherited from Female	18	28	30
Inherited from Male	23	51	50
Inherited from both parents		13[b]	99
Distorted (P<0.01)	10	8	38

[a] from 8 primer combinations

[b] These 13 bridge SNPs correspond in reality to 26 SNP markers as explained in Material and Methods.

Table IV.2 : Summary of genetic maps characteristics.

Map[a]	Mapped markers	Spanning (cM)	Average distance between adjacent markers (cM)	Number of LG	Unlinked markers
Female	174	1888.2	12.84	25 (6 doublets)	14
Male	224	2406.1	12.43	22 (3 doublets, 1 triplet)	12
Consensus	297	2188.7	10.27	24 (2 doublets)	15

[a] Female and male maps constructed with Mapmaker. Consensus map constructed with Joinmap

to refine QTL localization and to detect additional QTL potentially masked by previously detected ones.

QTL involved in epistatic interactions were searched among all marker pairs whose allelic class frequencies did not significantly (p=5%) deviate from the expected 1:1:1:1 segregation using the following linear model in R (The R development core team):

$$Y_{ijk} = \mu + M_i + N_j + (MN)_{ij} + \varepsilon_{ijk}$$

where μ is the grand mean, M and N are the fixed effects of markers M and N, (MN) is the fixed effect of the interaction between markers M and N, and ε is the residual error. We considered an epistatic interaction to be significant when the partial p-value of the (MN) interaction effect was below 0.05/b, where b is a Bonferonni correction factor corresponding to the number of tested marker pairs.

The overall percentage of phenotypic variation (global R^2) explained by all the detected QTL for a given trait by strain combination was computed with R (The R development core team) from a linear model where each additive or interactive QTL was represented by its nearest marker. This procedure allowed double checking the significance of each detected QTL.

C. Results

7. Marker screening and genetic map construction

A total of 750 markers were tested (**Table IV.S1** see annexe). The amplification success rate (percentage of amplified markers relative to markers tested) was on average 77.1%. This rate varied depending on the type of markers: SNP markers being more successful (94%) than AFLPs and SSRs (78%), because we designed the formers with the nucleotide sequence of the two *P. nigra* parents when SSR primers have been initially designed on *P. trichocarpa*. A successful selection of SSR markers associated to previous mapping in *P. nigra* induced as expected a higher success rate for the SSR markers selected from Gaudet et al. 2008 and Cervera et al. 2001), than for those coming from the IPGC web site (55%), underlining a low transferability of markers between *Populus* species (here from *P. trichocarpa* to *P. nigra*).

A total of 325 markers were polymorphic, 76 in the female parent, 124 in the male parent and 112 in both parents. The number of distorted markers inherited from the male parent was slightly higher (41) than from the female parent (29). AFLP markers showed the highest proportion of distortion (24%), with 7% having a $P\chi 2<0.0001$. Twenty-one percent of SSR markers were

Table IV.3 : Marker positions with discordant localization (non syntenic) between present work, genetic maps of (Gaudet et al. 2008) and (Cervera et al. 2001) and version 1.0, 2.0 and 3.0 of Nisqually-1 genome assembly.

Marker name	Genetic maps			Genome		
	Present study	Gaudet et al. (2008)	Cervera et al. (2001)	Nisqually-1 V1.0	Nisqually-1 V2.0	Nisqually-1 V3.0
ORPM_269	XI	VIII, XI	_	XI	scaffold 11	scaffold_1984&Chr11&Chr08
PMGC_451	XVII	XVII	C	scaffold_21039	scaffold 17	Chr17
PMGC_2550	IV	I	-	I	scaffold 1	scaffold_670&Chr01
WPMS_6	XVI	XVI	XII	no hit	scaffold 16	no hit
WPMS_4	VI	VI	-	VI	no hit	no hit
ORPM_16	XIII	XIII[a]	-	XIII	scaffold 30	no hit
PMGC_2832	XVII	NA	-	XVII	scaffold 9	Chr09
PMGC_2803	XVII	XVII	-	XVII	scaffold 9	Chr09
PMGC_2515	IV	IV	-	XIV	scaffold 14	no hit
GCPM_3893-1	V	-	-	VII	scaffold 7	scaffold_139&Chr07
ORPM_488	VI	-	-	XVIII	scaffold 18	Chr18
PMGC_2861	XI	-	-	scaffold_170	scaffold 9	Chr09
PMGC_433	XVI	-	XVI	scaffold_125	scaffold 19	scaffold_141&Chr19
scaffold_123_1_361	XVII	-	-	scaffold_123	scaffold 5	Chr17
bu890808	III	-	-	III	scaffold 22	Chr03
ORPM_26	VI	VI	-	VI	no hit	Chr06
scaffold_121_1_246	III	-	-	scaffold_121	scaffold 22	Chr03

[a] This marker was mapped at two loci on LG XIII.

distorted, with 11% having a Pχ2<0.0001, while SNP markers were in proportion the less distorted markers (7%, with only 4% having a Pχ2<0.0001).

In the two separate parental genetic maps, all linkage groups (LG) with at least 3 markers covered the 19 major scaffolds of poplar genome, except the chromosome 18 which was represented with only 2 markers in the LG XVIII. Of the 27 markers tested in this LG, only 7 were polymorphic underlining a low polymorphism level (**Table IV.S1** see annexe & **Figure IV.S1** see Chapter III, **Figure III.5**). The male map was longer than the female map because of a higher number of polymorphic markers in the male parent (**Table IV.2**). Thanks to 112 bridge markers (SSR and SNP), we elaborated the consensus map with the objective of reducing the number of linkage groups. A minimum LOD threshold of 6.0 was sufficient to get all the linkage groups, except for LG II and LG V and for LG I and LG VIII, for which the LOD threshold needed to be increased to 10.0 and 16.0 respectively, in order to split markers to their corresponding linkage group. The colinearity and synteny were respected between the separate parental maps and the consensus map except for regions harboring many distorted markers for which the marker colinearity was reversed. With the consensus map, we reduced the mean distance between neighboring markers but we obtained 24 linkage groups, more than in male map (22) and less than in female map (28) (**Table IV.2**). This distance reduction is probably due to the different algorithms used for marker ordering and distance calculation in Joinmap and mapmaker respectively.

A cluster of distorted markers, including all marker types (AFLP, SSR and SNP), was found on LG VI. Since each marker type corresponds to a single manipulation, we concluded that this distortion could not be attributed to a technical problem. Despite this distortion, the colinearity with the genome of Nisqually-1 *P. trichocarpa* and *P. nigra* maps was conserved.

In order to compare our map with the others genetic maps of *P. nigra*, the choice of SSR markers was based on the anterior SSR markers available on published *P. nigra* maps. We obtained 74 and 24 SSR markers in common with Gaudet et al. (2008) and Cervera et al. (2001) maps respectively. Additionally, 269 mapped markers are anchored on the Nisqually-1 *P. trichocarpa* genome (**Table IV.S1** see annexe page 212 & **Figure IV.S1** see chapter III, **Figure III.5**). These anchoring markers allowed analysing synteny and colinearity between *P. nigra* genome and *P. trichocarpa* genome. Globally, the synteny was conserved between our maps, published *P. nigra* maps and *P. trichocarpa* chromosomes. Indeed, only 17 marker discrepancies have been detected between our maps, both published *P. nigra* maps and the

three successive genome assemblies of Nisqually-1 *P. trichocarpa* genome (**Table IV.3**). Only four discordances between the three genetic maps were found, two of them being also divergent with one of several Nisqually-1 *P. trichocarpa* genome assemblies. Of the 9 synteny discordances between *P. nigra* maps and the last genome assembly, 3 cases corresponded to a default in the v1.0 and v2.0 genome assemblies (no hit) and 6 cases confirmed assignation to different chromosomes in genetic maps and genome assemblies. Both v1.0 and v3.0 assemblies were more consistent with genetic maps than v2.0 assembly.

The colinearity was respected between our maps and both (Gaudet et al. 2008) and (Cervera et al. 2001) map. A total of 12 colinearity discordances registered in the comparison between *P. nigra* genetic maps and *P. trichocarpa* genome sequence involving each time a difference in the position of one marker (**Figure IV.S1** see chapter III, **Figure III.5**)

8. Phenotypic variability for rust resistance epidemiological components

Three epidemiological components were measured on leaf disks after inoculation with two strains of *Mlp*, 98AG69 and 93JE3: Latent period (LP); uredinia number (UN); and average uredinia size (US) (**Figure IV.1**; **Table IV.S2**). The parents behaved differently depending on the *Mlp* strain used for inoculation. The male parent BDG was more resistant to 98AG69 strain than the female parent 71072-501. Indeed, BDG showed a longer LP, and fewer and smaller uredinia than 71072-501. For 93JE3 strain, there were no significant differences between the two parents for LP and UN (t-test= 0.631, p-value=0.533 and t-test=0.708, p-value=0.485 respectively) but male parent BDG showed slightly larger uredinia than 71072-501 (t-test= 2.183, p-value=0.04). The female parent 71072-501 demonstrated similar levels of resistance against the two strains except for UN where it showed more uredinia after 93JE3 infection than after 98AG69 infection due probably to a higher inoculum density (see material and methods).

The distributions of F1 genotypic means also differed depending on the strain both in terms of average performance and of expression of phenotypic variability (shape of histogram) (**Figure IV.1** and **Table IV.S2**). On average, the F1 progeny was more susceptible to 98AG69 than to 93JE3 infection with respect to the three epidemiological components. Whatever the epidemiological component, the F1 mean performance was quite close to the mean value of the female parent 71072-501 with 98AG69 strain and to those of both parents with 93JE3 strain, whereas a negative transgression with respect to the performance of the male parent BDG was observed in response to infection with 98AG69 strain.

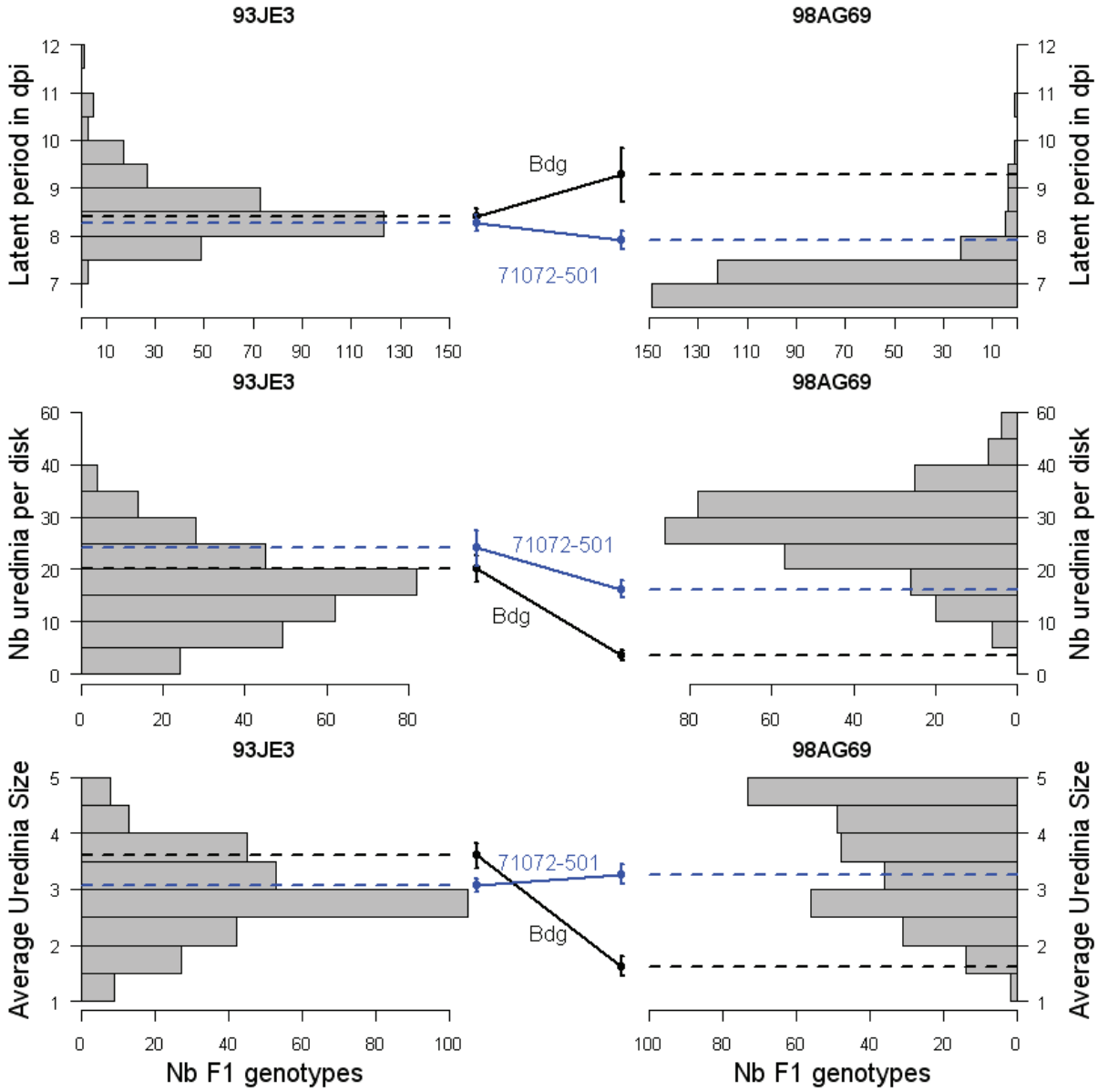


Figure IV.1 : Distributions of genotypic means for latent period (LP, days), uredinia number (UN) and uredinia size (US, 1-5 scale) following artificial inoculation of the F1 progeny with the two *Melampsora larici-populina* strains 93JE3 and 98AG69. Parental means with standard errors were positioned on graphs with horizontal dashed lines, blue for female 71072-501 parent and black for male BDG parent.

Significant genetic variability was observed for the three epidemiological traits measured with the two rust strain infections, which lead to medium to high broad sense heritability values (**Table IV.4**). Whatever the strain, the heritability estimates were higher for US than for LP and UN. They were also higher with 98AG69 than with 93JE3 for all traits except UN. These differences in heritabilities were explained by differences in the genetic variances, as the residual variances were almost the same after inoculation with the two strains for all three components.

To further analyse genotype x strain (GS) interactions, we constructed dedicated linear mixed models that allowed fitting two distinct residual variances (one for each *Mlp* strain infection). Compared to simpler linear mixed models that would fit a unique residual variance, these dedicated models had significantly better fits according to likelihood ratio tests (data not shown) and were thus selected for analysing the (GS) interactions. In such selected models, a significant genetic variance was detected for all traits, but the (GS) interaction variance was always much higher (**Table IV.4**). Interestingly, such interaction effects could not be attributed to few outliers as underlined by the low between strain rank correlations observed for all epidemiological components (**Figure IV.2**). In addition, it's worth mentioning that the (GS) interactions observed for US could not be explained by the ones detected for LP. For US, genetic variance and (GS) variance were higher and this was in agreement with the heritability values for this trait.

To investigate the genetic relationships between the three epidemiological components observed after inoculation with a given strain, rank correlations have been computed on the genotypic means. These correlations ranged between 0.34 and 0.66 in absolute value and their signs were consistent with the epidemiological expectation (**Figure IV.2**). Indeed, LP was negatively correlated with both UN and US, and UN and US were positively correlated. Nevertheless, the correlations between the different components were not very high in absolute value underlining that they are not completely redundant. This is especially the case for US and UN which displayed the weakest correlations in the two inoculation experiments.

Table IV.4 : Genetic variances with standard deviation (SD), broad sense heritabilities (H2) with standard error (SE) and relative contributions of genetic, genotype x strain (GS) interaction and residual variances in phenotypic variation in the F1 progeny for latent period (LP), uredinia number (UN) and size of uredinia (US) following artificial inoculation with the two *Melampsora larici-populina* strains 93JE3 and 98AG69.

	93JE3			98AG69			Multi-strain analysis		
	σ^2_G (SD)	σ^2_ε (SD)	H^2 (SE)	σ^2_G (SD)	σ^2_ε (SD)	H^2 (SE)	σ^2_G/σ^2_P %	σ^2_{GS}/σ^2_P %	$\sigma^2_R{}^a/\sigma^2_P$ %
LP (1/X²)	$2.28\ 10^{-6}$ ($2.5\ 10^{-7}$)	$3.89\ 10^{-6}$ ($1.6\ 10^{-7}$)	0.37 (0.03)	$2.81\ 10^{-6}$ ($2.7\ 10^{-7}$)	$3.92\ 10^{-6}$ ($1.2\ 10^{-7}$)	0.49 (0.03)	10.95 (3.14)	31.69 (3.25)	57.35 (2.08)
UN - square root (X)	$8.50\ 10^{-1}$ ($8.6\ 10^{-2}$)	$9.83\ 10^{-1}$ ($4.0\ 10^{-2}$)	0.46 (0.03)	$4.49\ 10^{-1}$ ($5.0\ 10^{-2}$)	$8.13\ 10^{-1}$ ($3.3\ 10^{-2}$)	0.36 (0.03)	12.10 (3.08)	30.21 (3.14)	57.69 (2.09)
US	$4.55\ 10^{-1}$ ($4.4\ 10^{-2}$)	$3.89\ 10^{-1}$ ($1.6\ 10^{-2}$)	0.54 (0.03)	$8.31\ 10^{-1}$ ($7.2\ 10^{-2}$)	$2.97\ 10^{-1}$ ($1.2\ 10^{-2}$)	0.74 (0.02)	15.40 (4.13)	49.83 (4.05)	34.77 (1.69)

[a] Residual: mean residual variances estimated for each strain.

Table IV.S2 : Parental and F1 progeny means with standard error (SE) for latent period (LP, days), uredinia number (UN) and size of uredinia (US, 1-5 scale) following artificial inoculation with the two *Melampsora larici-populina* strains 93JE3 and 98AG69.

	93JE3			98AG69		
	LP	**UN**	**US**	**LP**	**UN**	**US**
Male parent BDG	8.40 (0.18)	20.2 (2.5)	3.60 (0.21)	9.27 (0.57)	3.67 (0.9)	1.62 (0.18)
Female parent 71072-501	8.27 (0.17)	24.0 (3.4)	3.07 (0.12)	7.90 (0.20)	16.2 (1.7)	3.27 (0.18)
F1 Progeny	8.55 (0.03) n=308	16.54 (0.3) n=308	3 (0.02) n=308	7.22 (0.01) n=300	27.16 (0.3) n=301	3.62 (0.03) n=299
Genotypic mean range (min - max)	7.37 - 11.87	0.2 - 39.2	1 - 5	6.67 - 10.67	6 - 50	1.2 - 5

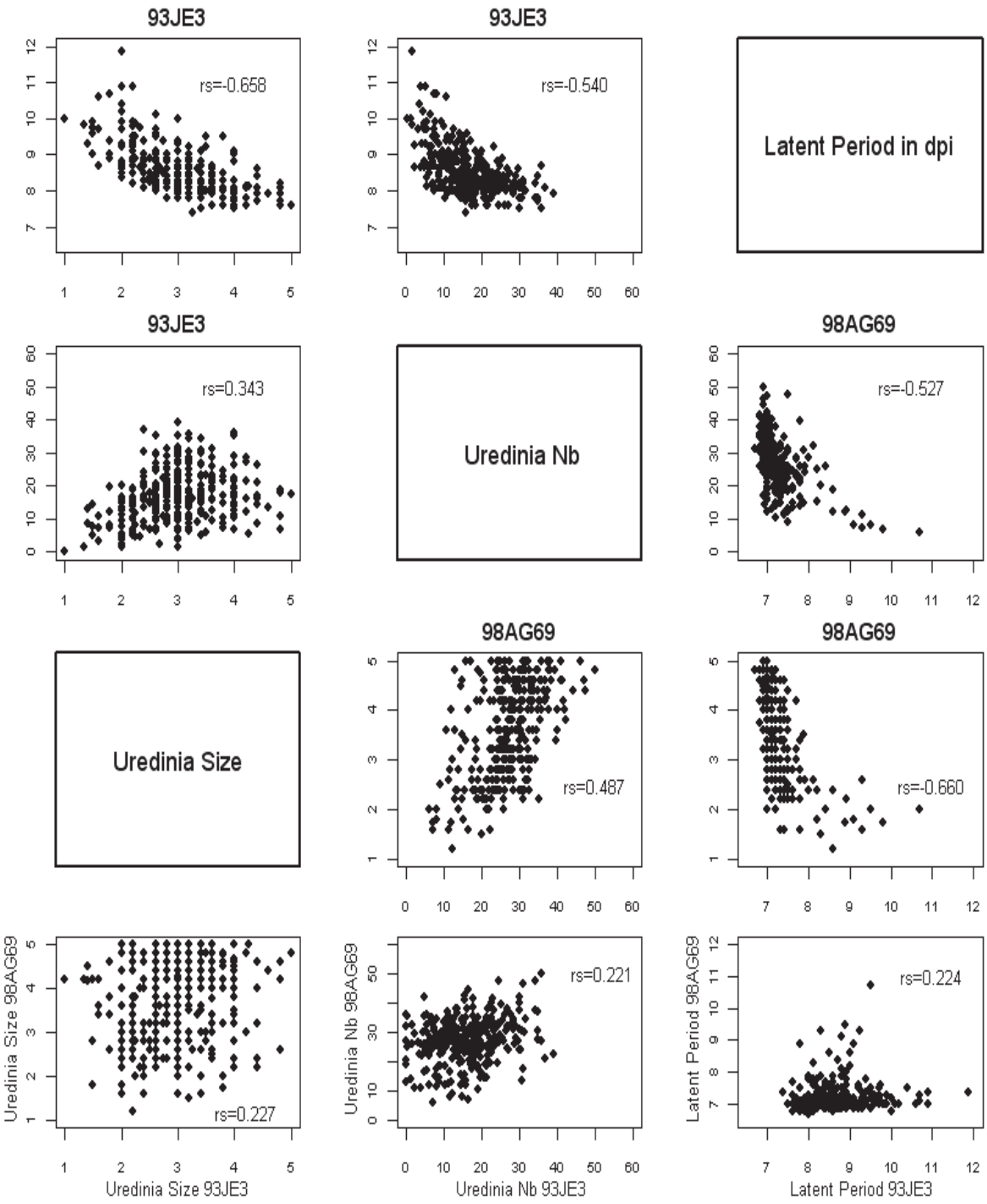


Figure IV.2 : Relationships between genotypic means for latent period (LP, days), uredinia number (UN) and size of uredinia (US, 1-5 scale) following artificial inoculation of the F1 progeny with the two *Melampsora larici-populina* strains 93JE3 and 98AG69. All Spearman rank correlations (rs) given inside each graph were significantly different from zero at 5% α-error risk. Not all possible correlations were represented: only intra-strain correlation between resistance components and inter-strain correlation for the same epidemiological component.

Table IV.5 : QTL detected on the consensus framework map using MQM method from Joinmap and significant markers using a linear model associated with the variation of three epidemiological traits: Latent period (LP); Uredinia number (UN); Uredinia size (US) following artificial inoculation with the two *Melampsora larici-populina* strains 93JE3 and 98AG69.

			MQM analysis[a]			Linear Model analysis[b]					
Strain	**Trait**	**LG**	**Position (cM)**	**CI (cM)**	**LOD**	**Marker(s)**	**Effect**	**P-val$_{ind}$**	**R^2_{ind}**	**P-val$_{glob}$**	**R^2_{glob}**
93JE3	LP	I	205.2	199 - 213	6.58	bu813833	M	$3.44.10^{-5}$	7.7	$6.99.10^{-5}$	22.9
		VIII[a]	33	1 - 53.4	2.75	WPMS_13	F	$3.97.10^{-3}$	4.9	$9.05.10^{-4}$	
		XII * XV	-	-	-	GCPM_1223-1 * GCPM_2743-1	F * M	$2.52.10^{-5}$	9.6	$2.55.10^{-4}$	
	UN	IV	1	0 - 12	7.86	PMGC_2881	M	$7.45.10^{-6}$	8.8	$3.20.10^{-6}$	15.2
		VII	33.9	31 - 34	4.76	PMGC_2140	F	$1.67.10^{-4}$	5.3	$1.71.10^{-4}$	
	US	I	201.2	197 - 205	4.89	bu813833	M	$5.94.10^{-4}$	5.8	$5.94.10^{-4}$	5.8
98AG69	LP	I	206.5	202 - 210	18.02	bu813833	M	$1.10.10^{-13}$	19.1	$1.10.10^{-13}$	19.1
	UN	I	209.5	204 - 213	20.92	bu813833	F & M	$1.80.10^{-17}$	23.7	$7.96.10^{-17}$	30.2
		VII	44.7	34.5 - 57.7	4.76	PMGC_2140	F	$5.73.10^{-5}$	6.0	$1.72.10^{-5}$	
	US	I	208.5	206 - 211	27.56	bu813833	M	$5.16.10^{-62}$	61.7	$1.23.10^{-59}$	66.0
		II	36.7	21 - 36.7	4.46	PMGC_667	M	0.58	0.7	$3.33.10^{-4}$	
		XVII	11.1	5 - 12.1	6	PMGC_2889	F	$7.45.10^{-4}$	5.5	$4.42.10^{-5}$	

[a] for linkage group VIII, the QTL is significant on female framework map only.

[b] for the linear model analysis, marker effects were tested individually and in interaction with each other, whatever the map they belong to. Individuals P values (P-val$_{ind}$) and R^2 (R^2_{ind}) were estimated from single marker models, while global P values (P-val$_{glob}$) and R^2 (R^2_{glob}) were estimated from global models *i.e.* including all significant markers together. This implies that P-val$_{glob}$ were partial P-values given the other markers.

9. QTL detection

For QTL detection, we used three genetic framework maps, two parental and one consensus framework maps, where we removed distorted and accessory markers. To build the framework consensus map, a minimum LOD threshold of 4.0 was chosen for the grouping step, except for LG XVII and LG VI, where the LOD value was increased to 6.0 to split these two linkage groups. The framework female map consisted in 36 linkage groups spanning 1490.8 cM, LG VI and LG XVIII were not present because the remaining markers in both groups were too distant to form a linkage group. The male framework consisted in 38 linkage groups spanning 1812 cM. The framework consensus was composed of 24 linkage groups spanning 1923.1 cM (Supplementary table 1).

In total, 11 QTL were detected explaining marginally between 2% and 60% of epidemiological trait variation (**Table IV.5**). For a given trait, between 1 and 3 QTL explained a total of 6% to 66% of the phenotypic variation.

A major QTL was detected on LG I for US_{98AG69} explaining 60% of phenotypic variation. It co-localized with QTL associated to all the other epidemiological traits except UN after infection with 93JE3 strain. All these 5 QTL were detected on both framework male and consensus maps. In addition, they all had similar allelic effects (male) but their magnitude was much stronger under 98AG69 infection (23.5 - 60%) than after inoculation with 93JE3 strain (6.9 - 9.6%; **Table IV.5**; **Figure IV.S2**).

Apart from LG I, other QTL were mapped on LGs II, IV, VII, VIII and XVII, explaining between 2 and 11% of trait variation (**Table IV.5**, **Figure IV.3**). On LG VII, the detected region harboured 2 QTL for UN detected after inoculation with both 98AG69 and 93JE3 strains. These QTL had similar allelic effect (female) and explained between 5.1 and 6.3 % of phenotypic variation (**Figure IV.S2**). Under 93JE3 infection only, another QTL was detected for UN on LG IV. This QTL explained 10.6 % of the phenotypic variance. The other QTL detected on LG II (US_{98AG69}), VIII (LP_{93JE3}) and XVII (US_{98AG69}) were quite small explaining between 2 and 4.1 % of the phenotypic variation.

To further investigate the genetic control of the epidemiological components, we carried out a genome scan for epistatic effects (**Table IV.5**). A significant interaction effect between loci was detected for LP under 93JE3 infection. This epistatic effect was found between 2 markers located on LG XII and LG XV and it explained 10% of the phenotypic variation. Including this epistatic effect, up to 24% of phenotypic variation observed for LP under 93JE3 infection could be explained, even if no strong QTL was detected.

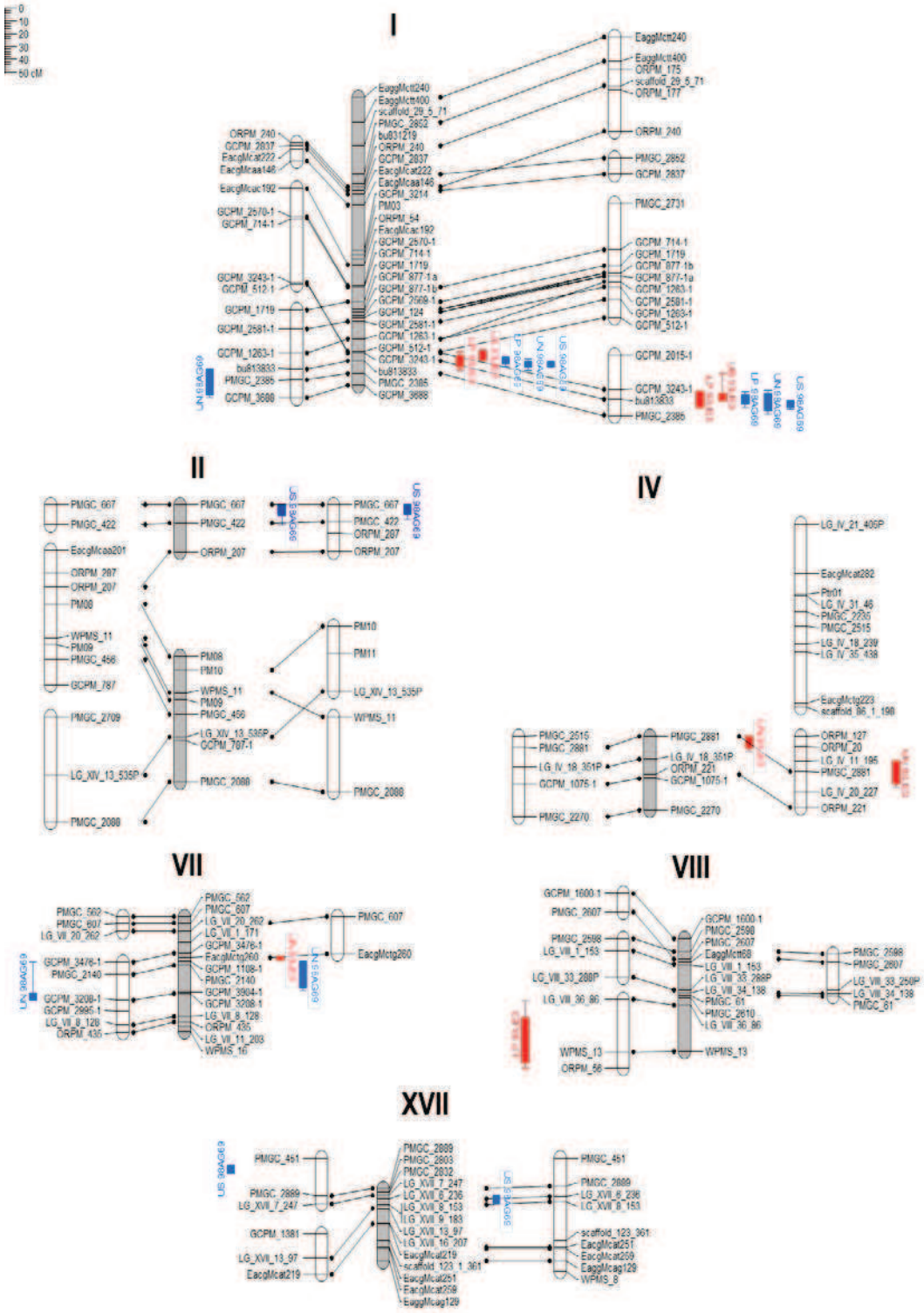


Figure IV.3 : Position of QTL controlling epidemiological components of resistance to two *Melampsora larici-populina* strains (98AG69 in blue and 93JE3 in red) on parental and consensus framework genetic maps constructed with the mapping population *P. nigra* 71072-501 x BDG. For each linkage group, the 71072-501 framework genetic map (left), consensus framework genetic map (middle, grey bars) and BDG framework genetic map (right) were aligned. QTL were represented by their confidence intervals (the length of the bar indicate the 2-LOD support interval and the line extension, the 1-LOD support interval). Trait abbreviations: LP, latent period; UN, uredinia number; US, uredinia size

D. Discussion

1. Added value of new *P. nigra* genetic maps

This work of identification of QTL has allowed the development of new genetic maps for the species *P. nigra*. Our strategy has been to both maximize *P. trichocarpa* genome anchoring with SSR and SNP markers and other *Populus* spp. genetic maps anchoring to allow comparative mapping. The alignment of these genetic maps to the *P. trichocarpa* genome and to other *P. nigra* genetic maps revealed few discrepancies, most of them being due to mis-assembling in the V2.0 or to duplicated regions across genome suggesting amplification of paralogs in *P. nigra* (e.g. GCPM_3893 on LG V in *P. nigra*, and on scaffold 7 on *P. trichocarpa* genome; ORPM 488 on LG VI in *P. nigra* and on scaffold 18 on *P. trichocarpa* genome). These genetic maps are still valid and can improve the assembly of genomes (e.g. assembly of scaffold 22 in scaffold 3 in *P. trichocarpa*) especially in new efforts will be devoted to increase marker density. Our genetic maps could thus also contribute to the expected genome assembly of a *de novo* sequence of *P. nigra* (M. Morgante and G. Zaina, pers. com.).

Some characteristics of the present genetic maps have a special additional interest to the resistance mapping aspects. First, we identified a large distorted region on LG VI, both in female and in male genetic maps (also in part monomorphic in female map; **Figure IV.S1** see chapter III, **Figure III.5**), suggesting the importance of this region for viability. Mortality rate after germination was negligible (9% with mortality occurring during the first three months in greenhouse), and thus rust or any other post-germination disease susceptibility could not be suspected to be associated with this distortion. Thus we postulate that the selection responsible for the observed segregation distortion has probably acted before the germination. Enrichment analysis of Gene Ontology terms in the distorted interval (data not shown) pointed out a gene controlling synthesis of sterols implicated in cell viability and plastid biogenesis in *Arabidopsis thaliana* (Babiychuk et al. 2008). The second noticeable feature of our genetic maps corresponded to the very low polymorphism observed on LG XVIII. Indeed, despite of a significant number of marker tested, this chromosome remains very poorly covered in our maps. This linkage group was also poorly covered in genetics maps reported before (family *P. deltoides* x *P. trichocarpa*; Jorge et al.2005). Most of the markers tested did not amplify. Surprisingly, the chromosome 18 is the "duplicated homolog" of a large part of the chromosome 6, a part corresponding to the distorted region mentioned before.

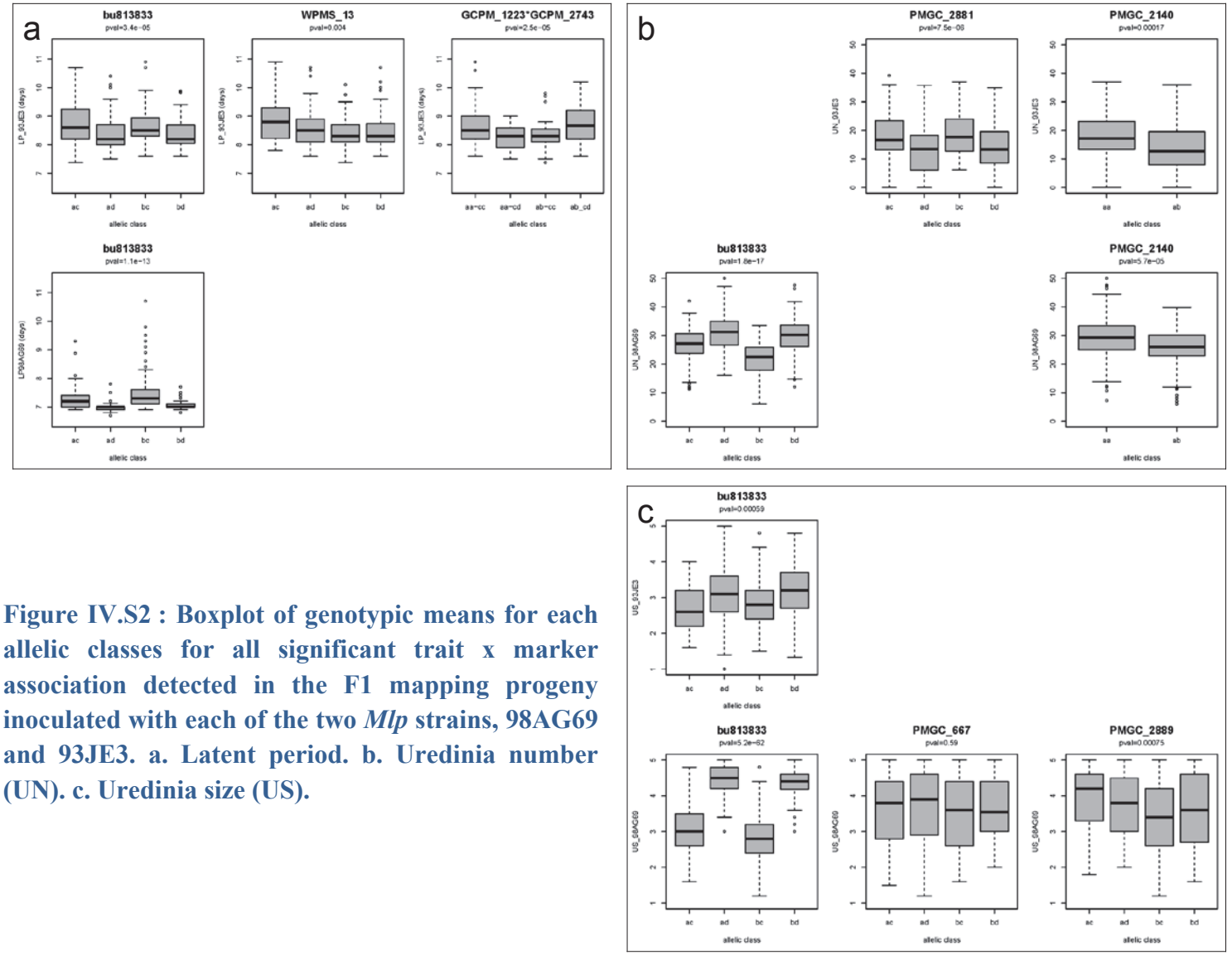


Figure IV.S2 : Boxplot of genotypic means for each allelic classes for all significant trait x marker association detected in the F1 mapping progeny inoculated with each of the two *Mlp* strains, 98AG69 and 93JE3. a. Latent period. b. Uredinia number (UN). c. Uredinia size (US).

2. First identification of QTL associated to *Mlp* rust resistance in *P. nigra*

This is the first reported study on the genetic architecture of *P. nigra* partial resistance to *Mlp,* one *Melampsora* species that have co-evolved with this pathogen in its natural range in Eurasia. As field resistance can be influenced by the large diversity of *Mlp* populations (Pinon and Frey 2005; Xhaard et al. 2011), artificial inoculation tests using *Mlp* single-uredinial strains allowed more precise dissection of genetic control of partial resistance. After inoculation with two *Mlp* strains, we detected a total of 8 genomic regions distributed on 6 different linkage groups and explaining between 2% and 60% of phenotypic variation for three epidemiological components of partial resistance. The genetic control of partial resistance observed here in that multiple loci with minor to major effects were involved individually or in interaction and that there were specific interactions between particular QTL and different strains was in agreement with available references on *Salicaceae* leaf rust resistance. Previous QTL detection studies were dedicated to identification of genetic and molecular basis of poplar resistance to *Mlp* in interspecific F1 families segregating for exapted resistance (*sensu* Gould and Vrba, 1982, referred by (Newcombe 1998) inherited from *P. deltoides* and/or *P. trichocarpa* (Cervera et al. 1996; Zhang et al. 2001; Jorge et al. 2005; Bresson et al. 2011). In a F1 *P. deltoides x P. trichocarpa pedigree*, Jorge et al. (2005) detected a total of 19 QTL on three linkage groups of the female parent and one linkage group of the male parent for the same three epidemiological components after inoculation with seven different *Mlp* strains. However, for 98AG69, the unique strain in common with the present work, two and four different genomic regions were detected in the interspecific and the present *P. nigra* intraspecific pedigrees respectively. The cluster of 3 QTL detected on *P. trichocarpa* LG XIX explained between 24.8 and 81% of phenotypic variation for LP, UN and US when the cluster of 2 QTL detected on *P. deltoides* LG XIX explained less than 10% of phenotypic variation for LP and UN. In the pathosystem involving *Salix* host species and *Melampsora larici-epitea (Mle)*, Samils et al. (2011) also reported QTL detection in a *S. viminalis* intraspecific family with coadapted potential sources of resistance and in an interspecific family where an exapted resistance inherited from *S. schwerinii* is segregating. Thanks to artificial inoculation with three *Mle* strains, they detected on three linkage groups of the interspecific genetic map, 6 different genomic regions associated to LP, UN and US from which four were mainly explained by *S. schwerinii* parental moderate to high effects. With the same three strains, they detected on 6 different linkage groups of the intraspecific consensus map, 6 genomic regions with moderate effects associated to LP, UN and US. Considering

comparable number of tested *Mlp* strains, size of mapping pedigrees , average marker density on genetic maps and phenotypic variation for rust resistance, we thus concluded that the genetic architecture of partial resistance to *Mlp* is more complex in *P. nigra,* the coadapted host species than in interspecific genetic backgrounds involving exapted resistance factors in agreement with other results on a similar pathosystem involving a close related genus (*Salix*). Most realistic models of plant-pathogen coevolution are postulating today a 'gene-for-genome' hypothesis where the genome-wide dynamics of resistance sources modulate plant response to pathogen attacks, not simply allelic change at individual loci (Mitchell-Olds and Bergelson 2000).

3. A strong genotype by strain interaction

A strong genotype by strain interaction which concerned all resistance components was observed for the male parent BDG and in the progeny. This interaction combined both rank and scale changes of genotypic performances which resulted both in different sets of QTL detected and changes of the effect magnitude of the overlapping QTL. Differences in expression of genetic variation for rust resistance could be partly explained by the difference in the aggressiveness of the two *Mlp* strains, 98AG69 being more aggressive on the studied family than 93JE3 as underlined by shorter LP and larger classes of genotypes with large US.

The most pronounced differences between the two inoculation tests concerned US for which the rank changes were important and the genotypic classes with larger values after inoculation with strain 98AG69 represented a second mode in the distribution. Interestingly, the male parent BDG changed significantly rank with parent female 72072-501 for the same epidemiological component. Similar patterns of genotype x strain interactions for LP and less pronounced phenotypic plasticity for UN and US have already been described in a collection of *P. nigra* clones sampled in natural populations and tested with three *Mlp* strains (Legionnet et al. 1999). Regarding poplar partial resistance to *Mlp*, moderate to high strain-specificity was up to now the rule rather than the exception. Specificity with one or few *Mlp* strains of different exapted resistance factors inherited from *P. deltoides* has been demonstrated in interspecific *P. deltoides x P. trichocarpa* F1 clones (Dowkiw and Bastien 2007, 2004). Specificity of R_{US}, a major exapted quantitative resistance factor involved in US control and inherited from *P. trichocarpa* was more recently revealed with identification of R_{US}-defeating isolates in natural *Mlp* populations (Dowkiw et al. 2010). In *Salix-Mle* pathosystem, both Samils et al. (2011) and (Hanley et al. 2011) observed on different pedigrees higher correlations between genotypic performances evaluated with four and two *Mle* strains respectively but a different pattern of host-pathotype interaction was also observed with one particular strain more genetically distant

(Samils et al. 2011). As natural inoculum is a mixture of a large number of strains, these results highlighted the complex genetic control of field resistance in *P. nigra* (Legionnet et al. 1999) and the need to perform artificial inoculation tests with a larger number of *Mlp* strains for the identification of a more comprehensive list of resistance factors of interest in breeding programs.

4. A combination of major and minor resistance loci revealed in *P. nigra*

QTL detected in this study can be classified in three categories. First, as expected from the strong genotype x strain interaction observed, only two QTL were detected for the same epidemiological component evaluated with the two strains. One minor QTL explaining 5.3% and 6% of phenotypic variation for UN is located on LG VII with the greatest allelic effect coming from the female parent. Despite a variation of magnitude of its effects, the second locus located on LG I is of particular interest for several reasons. This major locus had significant effects on LP with both strains and on UN and US with 98AG68 strain. The greatest allelic effect at this locus came from the most resistant parent BDG. Thanks to the strong colinearity and synteny between *Populus* and *Salix* genomes, we could determine that this QTL was located in the same chromosomal region than the major QTL probably inherited from *Salix schwerinii* and detected by Samils et al. (2011) and Hanley et al. (2011) in two different pedigrees on both field and laboratory resistances. Kohler et al.(2008) identified clusters of putative nucleotide-binding site leucine-rich repeat (NB-LRR) genes on *P. trichocarpa* genome (Kohler et al. 2008), one of them being in the region harboring the major *P. nigra* and *Salix* QTL on LG I, and consisting in 25 NB-LRR genes. R_1 and R_{US}, two loci involved in poplar resistance to *Mlp* with major but variable effects on epidemiological components both belonged to a genomic region on chromosome 19, also rich in NB-LRR and serine threonine kinase (STK) genes (Jorge et al. 2005; Bresson et al. 2011). Although a challenge, the identification of the polymorphism underlying this major QTL could provide valuable information about nature of coadapted partial resistance in *P. nigra* and its relationship with R-genes. Before initiating cloning *via* positional strategies, additional fine mapping and phenotyping evaluation will be required with more markers anchored on poplar genome, larger size of the mapping family and replicated inoculation tests. Next-generation whole-genome resequencing will offer soon an alternative to positional cloning by sequence analysis of contrasted bulks constructed from extreme phenotypes in the pedigree (Hillier et al. 2008).

Second, four strain-specific QTL distributed over four linkage groups (II, IV, VII and XVII) were detected with one or other of the two strains and for only one epidemiological components. They all presented small to moderate effects with between 2% and 11% of trait variation

explained and the greatest favorable effects came either from the male or female parent. Comparative mapping using SSR marker genome position positioned approximately the *MCX3* resistance locus to *Melampsora. x columbiana* on the sub-telomeric region of the chromosome 4 whereas the QTL detected in the present study was positioned in the middle of the chromosome (Stirling et al. 2001; Yin et al. 2004a). Despite of this chromosome level co-location, we could not speculate on a common genetic control of resistance to different *Melampsora* spp.

Third, the large progeny size used in the present study allowed performing a systematic scan of interactions between pairs of loci which lead twelve significant epistatic effects for the three epidemiological components (**Figure IV.S2**). The epistatic effect involving two loci located on LG XII and LG XV respectively is of particular interest. The double homozygous or the double heterozygous genotypes were more susceptible than the genotypes having homozygous/heterozygous genotypes. Interestingly, these two linkage groups correspond to a chromosome "couple" that has rose from the whole-genome *Salicoid* duplication (Tuskan et al. 2006). Epistasis between paralog genes has been studied and reported in model species (Musso et al. 2008) but the advantage detected here has to be confirmed by mapping supplemental markers to both linkage groups before speculating on any molecular mechanism.

In order to extend/generalized further this first identification of QTL for resistance to *Mlp* in *P. nigra*, more *Mlp* strains need to be tested to determine the spectrum of resistance of the QTL identified. Additionally, we need to improve the marker density to narrow QTL intervals and lists of positional candidate genes in which polymorphism will be analyzed. As the mapping pedigree used is part of a 5x4 factorial mating design already evaluated for both field and laboratory resistance, extension of QTL detection to this material will soon allow both validation of QTL detected in the present study and identification of other loci of interest.

5. Phenotypic dissection of partial resistance to *Mlp* in *P. nigra*

The plant response to artificial inoculation has been decomposed in three epidemiological components showing distributions with different shapes and range of variation confirming the interest of such decomposition to analyze genotype x strain interactions. Latent period and spore production indirectly assessed by average uredinia size in this experiment have been regarded as the most important variables for quantifying both pathogen aggressiveness and partial resistance in other plant models (Rime et al. 2005; Caffier et al. 2010). Genetic variability for these two components in poplar were not affected by inoculum pressure which had only a scale effect and did not change either genotype ranking for uredinia number (Dowkiw et al. 2003). In the two inoculation tests, the correlations between the three components were moderate to high, the

highest being observed between LP and US as previously described in similar studies (Samils et al. 2011; Dowkiw and Bastien 2004). With left-triangular shaped relationship, this correlation was probably overestimated by particular performance of the latest sporulating genotypes as for latent period exceeding ten days, the size of uredinia evaluated at 14 days after infection was significantly restrained. When latent period was less than 10 days with strain 93JE3 and less than 8.5 days with strain 98AG69, correlations between LP, UN and US were obviously lower and let expect small overlap between genetic controls of the three parameters. These results confirmed previous observations from Legionnet et al. (1999) highlighting the possibility to identify in *P. nigra* a diverse complement of resistance genes which could be successfully combined in elite genotypes.

E. Conclusion

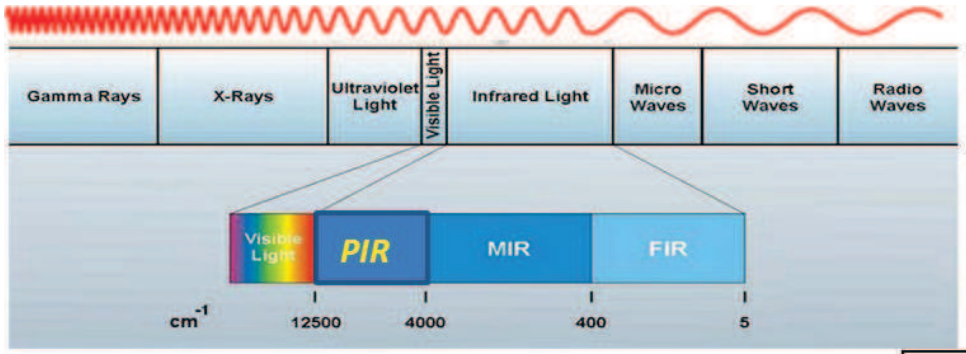
Durable resistance to *Mlp* natural attacks in poplar plantations has remained a key target for genetic improvement programs in Europe. Most of interspecific hybrid clones deployed today have *P. nigra* as one of the parental components. Failures of past breeding strategies taking mostly value of complete resistances inherited from *P. deltoides* proved that it is essential to devote more efforts on prospection and combination of partial resistances present in *P. nigra* natural populations which co-evolve with *Mlp* populations. Although the efficacy of this strategy to develop durable resistance in *P. deltoides x P. nigra* genetic background is still to be proven in practical terms, the first insights in genetic architecture of partial resistance to *Mlp* from the present study are encouraging.

Acknowledgments

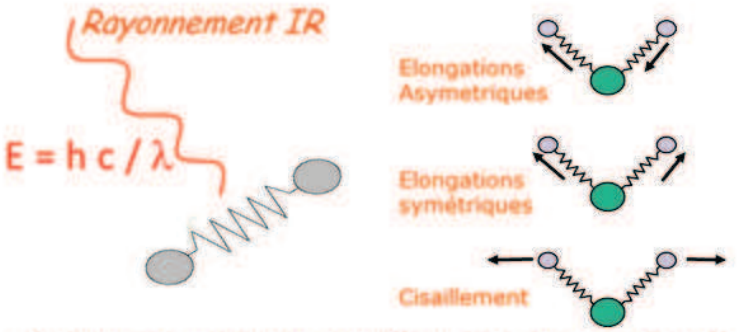
R. El Malki was supported by a Ph.D. grant in the behalf of Futurol project, and authors thanks OSEO Innovation who participates in funding this project. This research was also supported by the European Commission through the directorate General Research within sixth framework programme (Network of Excellence EVOLTREE -contract n° 16322), by INRA (AIP Bioressources), and by Bureau des Ressources Génétiques. Authors thank Marc Villar for the initial creation of the plant material from the 5x4 factorial mating design.

Chapitre V

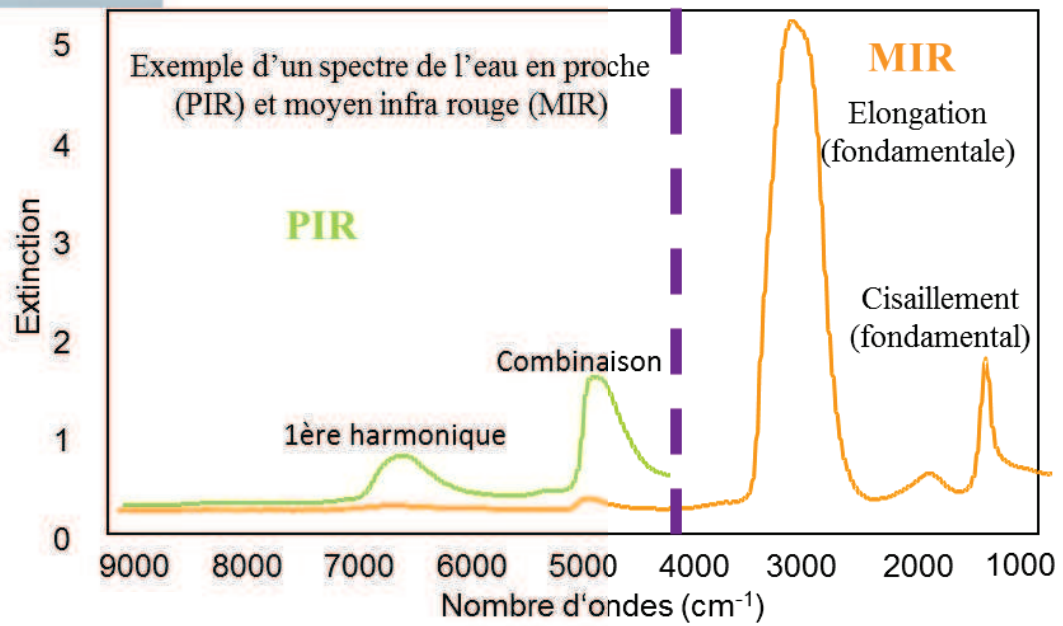
- Cartographie de QTL associés à la qualité de la biomasse



Les liaisons chimiques sont capables d'absorber les rayonnements correspondants à certaines fréquences particulières.
C'est cette propriété des liaisons chimiques de la matière organique que va utiliser la spectroscopie dans le proche infrarouge (PIR) pour établir un lien entre l'absorption de la lumière et la composition de l'échantillon.



Quelques exemples des vibrations entre atomes détectés en PIR



Dans la zone du proche infrarouge, les **absorptions** ne sont pas dues aux vibrations fondamentales des molécules (détectées en MIR), mais aux **vibrations harmoniques** et aux **vibrations de combinaisons**.

Figure V.1 : Principe de la spectroscopie en proche et moyen infrarouge

Chapitre V: Cartographie de QTL associés à la qualité de la biomasse

A. Méthodes d'évaluation directes et indirectes des composants chimiques du bois

Malgré son importance considérable dans le procédé industriel, la qualité de la ressource ligno-cellulosique est restée assez ignorée jusqu'à présent dans les programmes d'amélioration génétique. Son intégration en tant que critère de sélection nécessite de disposer de méthodes d'évaluation fiables et à haut débit afin de pouvoir cribler de nombreux candidats à la sélection. Le développement de méthodes de phénotypage fiables, faciles à mettre en œuvre et peu coûteuses est donc indispensable, puisque les méthodes de référence sont souvent coûteuses et laborieuses et ne permettent de traiter au maximum que quelques dizaines d'échantillons par jour. Plusieurs méthodes ont également été développées pour évaluer les propriétés du bois de façon non destructive. Ainsi, on est capable d'estimer la densité du bois et son hétérogénéité de façon rapide et précise sur des carottes de bois de petite taille dont le prélèvement ne perturbe pas la croissance de l'arbre. De même, les mesures de morphologie des fibres autrefois mesurées avec la micrographie (technique très lourde à mettre en place) se font actuellement avec des appareils d'analyse d'image numérique (revue par Wessels et al. (2011)).

Pour la détermination de la composition chimique du bois, il existe des méthodes directes et indirectes. La spectrométrie de masse est une méthode directe qui a été proposée pour analyser le taux de monomères de sucres C6 et C5 hydrolysés, le taux de lignine et le ratio S/G de la lignine (Novaes et al. 2009; Sykes et al. 2009; Yin et al. 2010 ; Guerra et al. 2013). Cette méthode a été utilisée pour estimer le taux de lignine de 800 génotypes de peuplier. Elle permet de détecter après pyrolyse différents pics assignés aux différentes composantes des lignines à savoir les unités H, S et G. En sommant la totalité de ces unités, on obtient le pourcentage de la lignine. On peut aussi facilement calculer le ratio S/G (Sykes et al. 2009). La spectrométrie de masse a été proposée aussi pour discriminer des mutants de peuplier qui sont affectés pour leur taux de lignine et de sucres libérés après pyrolyse (Meier et al. 2005).

La Spectroscopie dans le Proche ou Moyen Infra Rouge (SPIR ou SMIR) est quant à elle une méthode indirecte pour prédire la qualité du bois. Elle s'est développée il y a quelques années seulement pour caractériser le bois (Bailleres et al. 2002; Wessels et al. 2011). Il s'agit d'une technique analytique basée sur le principe d'absorption des rayonnements (infrarouges) par la matière organique. Cette absorption liée à la composition chimique des échantillons permet d'estimer cette dernière par la simple acquisition de spectres d'absorption de

Tableau V.1 : Gamme de variation de la teneur en lignines, du ratio S/G, des teneurs en sucres à cinq (C5) et six atomes de carbones (C6) avec leurs valeurs d'héritabilités au sens large (H2) chez différentes espèces et hybrides de peuplier. *P. trichocarpa :* T ; *P. deltoides* : D ; *P. nigra* : N. Spectroscopie de masse après pyrolyse : pyMBMS. N- sans azote, N+ avec 25 mM d'azote.

Espèces	Fond génétique	Nombre de clones	Age / Modalité	Méthode de phénotypage	Caractère								Source
					Lignines	H^2	Ratio S/G	H^2	C6	H^2	C5	H^2	
T	Population d'association	1100	inconnu	pyMBMS	15,7 - 27,9		1,0 – 3,0						Studer et al. (2011)
		448	inconnu	pyMBMS	15,8 – 27,5		1,2 – 2,4						Wegrzyn et al. (2010)
N	17 familles en pollinisation libre	34-36 / famille, 599 au total	2 ans	pyMBMS	23,6 - 25,2	0,58	1,6 - 1,8	0,70	32,2 - 34,5	0,46	29,2 - 31	0,48	Guerra et al. (2013)
(T x D) x (T x D)	F2	95	4 ans	Analyse chimique	23,4 - 32,1								Zhou et al. (2011)
		5	8 ans	pyMBMS	22,7 - 25,8		1,8 - 2,3						Davison et al. (2006)
		~ 290	inconnu	pyMBMS	24,3 (1,06)*		2,01 (0,16)*						Yin et al. (2010)
(T x D) x D	Pseudo rétrocroise-ment	396	2,5 mois / N-	pyMBMS	21,69 (0,06)*	0,23	0,99 (0,004)*	0,38	32,89 (0,07)*	0,17	25,68 (0,04)*	0,163	Novaes et al. (2009)
			2,5 mois / N+		17,38 (0,06)*		0,92 (0,004)*		35,54 (0,14)*		27,02 (0,08)*		

* les moyennes et écart-types correspondants sont indiqués car les gammes de variations ne sont pas disponibles.

l'échantillon (**Figure V.1**). La SPIR et la SMIR nécessitent toutes deux une phase d'étalonnage ou de calibration basée sur des mesures de référence obtenues au laboratoire et l'établissement des modèles mathématiques qui permettront de relier le spectre infrarouge aux résultats de ces mesures. Elle a été utilisée chez différentes espèces forestières pour estimer le taux de lignines (Bailleres et al. 2002 ; Yeh et al. 2004 ; Hou and Li 2011 ; Zhou et al. 2011), de cellulose et le rendement de la saccharification (Hou and Li 2011), et le ratio S/G (Robinson and Mansfield 2009) ou encore le pouvoir calorifique de la biomasse (Zhou et al. 2011). Les modèles rapportés sont de bonne qualité avec des R^2 de validation croisée compris entre 0,67 et 0,99 selon le caractère, l'espèce et la méthode d'analyse.

Le développement de ces méthodes de phénotypage des propriétés du bois à haut-débit a permis récemment l'exploration de la diversité génétique de ces caractères chez diverses espèces forestières comme l'eucalyptus (Thumma et al. 2010), le saule (Brereton et al. 2010; Ray et al. 2012), le pin *taeda* (Gonzalez-Martinez et al. 2007), le pin maritime (Lepoittevin et al. 2012) et le peuplier (Davison et al. 2006 ; Novaes et al. 2009 ; Yin et al. 2010 ; Wegrzyn et al. 2010 ; Studer et al. 2011; Guerra et al. 2013). Le haut débit de ces méthodes indirectes se caractérise par environ 42 échantillons par heure en pyrolyse suivie de spectrométrie de masse (Sykes et al. 2009) et 12 échantillons par heure en SPIR (K. Ader comm. pers.). Ces débits sont à comparer à ceux des méthodes chimiques de référence qui par exemple sont de 4 à 8 échantillons par jour pour une évaluation du taux de lignine Klason (K. Ader, comm. pers.).

B. Analyse du déterminisme génétique et moléculaire de la qualité de la biomasse : état de l'art chez le peuplier

Chez le peuplier, diverses études ont été menées pour étudier la variabilité génétique au sein de différentes espèces. Le peuplier représente un important réservoir de variabilité phénotypique pour la composition chimique du bois. Le contenu en lignine varie entre 15,8 et 27,8%, le ratio S/G varie entre 0,91 et 3, les sucres C6 varient entre 32,2 et 35,5% et le pourcentage des sucres C5 varie entre 25,6 et 27% (**Tableau V.1**). Peu d'études ont été effectuées pour mesurer la part de la variance génétique totale ou additive dans la variabilité phénotypique observée. Novaes et al. (2009) trouvent des valeurs d'héritabilité au sens large faibles avec un maximum de 0,38 pour le ratio S/G dans une famille F1 d'un peuplier hybride (*P. trichocarpa* x *P. deltoides*) x *P. deltoides*, alors que Guerra et al.(2013) trouvent des valeurs d'héritabilités au sens large moyennes à élevées (**Tableau V.1**) dans 17 familles de demi-frères chez *P. nigra*.

Le déterminisme génétique de ces caractères d'intérêt a jusqu'à présent été principalement abordé avec deux approches : (i) la détection de QTL dans des pédigrées de cartographie génétique intra ou interspécifique ; (ii) la génétique d'association dans des populations plus au moins complexes et dans des gènes candidats définis au préalable pour leur implication dans la voie de biosynthèse des lignines, de la cellulose et de l'hémicellulose. Si la première approche a généralement une faible résolution et ne permet d'identifier que les polymorphismes qui ségrégent dans le pédigrée étudié, elle présente l'avantage de pouvoir être effectuée sur tout le génome. À l'inverse, l'approche de génétique d'association dans des gènes candidats présente l'avantage d'être potentiellement bien plus précise et, selon la complexité du matériel, de pouvoir cribler un nombre important de polymorphismes, mais l'inconvénient est de ne pas être exhaustif sur tout le génome. Ce dernier point pourrait être compensé par une génétique d'association génome entier qui balaye les associations sur tout le génome. Mais, compte tenu de la faible étendue du déséquilibre de liaison chez le peuplier cela nécessiterait un nombre de marqueurs extrêmement important (plusieurs dizaines voire centaines de milliers).

Ainsi, le choix des candidats apparait crucial pour le succès de l'approche de génétique d'association. Les gènes candidats sont généralement expressionnels ou fonctionnels. Pour les propriétés chimiques du bois, la transformation génétique a permis de valider le rôle fonctionnel de plusieurs gènes dans le contrôle du taux de lignine. Des mutants transgéniques des gènes CAOMT (CaffeicAcid O-MethylTransferase), 4-coumarate : coenzyme A ligase et CCR (Cinnamoyl-CoA Reductase) impliqués dans la voie de biosynthèse des lignines ont été obtenus par interférence ARN sens et antisens (Tsai et al. 1998; Hu et al. 1999; Leple et al. 2007). Une réduction des transcrits de 3 à 4 % dans les mutants CCR provoque une réduction de 50% du taux de lignine total chez le peuplier transgénique (Leple et al. 2007). La baisse de la teneur en lignine chez la luzerne transgénique a permis d'améliorer le taux de sucres hydrolysés (Chen and Dixon 2007). Toutefois, les efforts récents de la physiologie moléculaire boostés par les avancées technologiques et surtout les nouvelles générations de séquençage (Metzker 2010) ont permis de générer des listes de candidats de plus en plus longues qu'il convient d'affiner pour mettre en œuvre les approches d'associations avec succès. Dans cet objectif, une stratégie couplant détection de QTL et génétique d'association peut être envisagée. En effet, la cartographie de QTL peut représenter le point de départ pour une cartographie fine via l'analyse des gènes sous-jacents du QTL en se basant sur une séquence de référence et ainsi définir des candidats positionnels.

Chez le peuplier, l'étude des déterminismes génétiques de la composition chimique du bois a été entreprise avec ces deux approches séparément. La génétique d'association dans des gènes

Tableau V.2 : Synthèse bibliographique des associations significatives de SNP appartenant à des gènes candidats avec la teneur en lignines, le ratio S/G, la teneur en sucres à six carbones (C6) chez *P. trichocarpa* (Wegrzyn et al. 2010) et chez *P. nigra* (Guerra et al. 2013). Le chiffre indique le nombre de SNP significativement associés et le pourcentage de la variation phénotypique expliquée par ces SNP est indiqué entre parenthèse.

Produit du gène	Gène	*P. trichocarpa*			*P. nigra*	
		Lignines	Ratio S/G	C6	Lignines	Ratio S/G
4-Coumarate:CoA ligase	4CL1 4CL2 4CL3 4CL5					
Coumarate 3-hydroxylase	C3H1 C3H3					
Coumarate 4-hydroxylase	C4H1 C4H2	1 (1,87) 1 (3,84)		2 (1,37 - 3,73)		
Cinnamyl alcohol dehydrogenase	CAD		1 (3,22)	1 (1,75)		
Cinnamoyl-CoA reductase	CCR	1 (1,19)				
Cellulose synthase	CesA1A CesA1B CesA2A CesA2B CesA3A	1 (1,63) 3 (1,63 - 2,27) 1 (1,72)		3 (1,38 - 2,18) 2 (1,31) 1 (1,57) 1 (1,35)		
Caffeoyl CoA O-methyltransferase	CoAOMT1 CoAOMT2			1 (2,72)		2 (~ 5)
Caffeate O-methyltransferase	COMT1 COMT2			1 (1,4)		
Ferulate 5-hydroxylase	F5H1 F5H2 F5H3 F5H4					
Glycine decarboxylase complex, H	gdcH1					
Glycine decarboxylase complex, T	gdcT2					
Hydroxcinnamoyl-CoA transferase	HCT1 HCT6	1 (1,59) 1 (2,17)		1 (1,22)		
Cellulase	KOR1					
Laccase	LAC1A LAC2 LAC90A			2 (1,11 - 1,84)		
Phenylalanine ammonia-lyase	PAL2 PAL4 PAL5			1 (1,35)		
S-Adenosylmethionine synthetase	SAM1			1 (1,62)		
Serine hydroxymethyl transferase	SHMT1 SHMT3 SHMT6					
Sucrose synthase	SUSY1	3 (1,52 - 2,68)		6 (1,18 - 2,07)		
α-Tubulin	TUA1 TUA5					
β-Tubulin	TUB15 TUB16 TUB9				4 (0,1 - 14,9)	

candidats de la voie de biosynthèse de la lignine et de la cellulose, a été récemment mise en œuvre chez *P. trichocarpa* (Wegrzyn et al. 2010) et chez *P. nigra* (Guerra et al. 2013). Chez *P. trichocarpa*, un total de 37 SNP appartenant à 16 gènes sont significativement associés à la teneur en lignines, au ratio S/G et au taux de sucres à six carbones. Ces associations expliquent entre 1,19 et 3,84% de la variation phénotypique de ces caractères (**Tableau V.2**). Chez *P. nigra*, peu de SNP significatifs ont été détectés. Au total, 6 SNP appartenant à 2 gènes sont associés au ratio S/G et à la lignine et expliquent 10 et 5% de la variation phénotypique pour ces deux caractères respectivement (**Tableau V.2**). Plusieurs études de détection de QTL dans des pédigrées pour des caractères de qualité du bois ont aussi été développées chez le peuplier (Novaes et al. 2009; Yin et al. 2010) et chez le saule (Brereton et al. 2010) qui appartiennent tous deux à la famille des Salicacées. Les génomes de ces deux genres sont caractérisés par une forte synténie et un degré élevé de colinéarité (Hanley et al. 2007 ; Berlin et al. 2010). Yin et al. (2010) ont analysé la composition chimique des tiges dans une famille F2 (*P. trichocarpa* x *P. deltoides*) x (*P. trichocarpa* x *P. deltoides*). Quatre QTL pour la teneur en lignines ont été détectés sur 3 groupes de liaison expliquant entre 4,6 et 7,6% de la variation phénotypique et 5 QTL pour le ratio S/G localisés sur 5 groupes de liaison et expliquant entre 5,2 et 12,5% de la variation phénotypique. Parmi ces QTL, 2 co-localisaient sur le groupe de liaison VI (Yin et al. 2010). Dans un autre pédigrée issu d'hybridation interspécifique (*P. trichocarpa* x *P. deltoides*) x *P. deltoides*, Novaes et al. (2009) ont cartographié des QTL associés aux teneurs en sucres C5 et C6, au taux de lignine et au ratio S/G. 4 QTL associés aux teneurs en sucres C5 localisés sur 4 groupes de liaison différents expliquaient entre 5,03% et 8,79 % de la variation phénotypique. Quatre QTLs localisés sur 3 groupes de liaison différents expliquaient entre 3,85 et 11,24% de la variabilité phénotypique pour la teneur en sucres C6. Un QTL associé au taux de lignine expliquait 9,61% de la variation phénotypique pour ce caractère et co-localisait sur le groupe de liaison XII avec le QTL identifié pour la teneur en sucres C6. Enfin, 3 QTLs localisés sur 3 groupes de liaison différents expliquaient entre 3,88 et 5,03 % de la variation phénotypique pour le ratio S/G. Chez le saule, des QTL pour le rendement en éthanol ont été détectés sur 4 groupes de liaison différents. Ces QTL expliquaient entre 10,52 et 18,5% de la variation phénotypique (Brereton et al. 2010). En conclusion, les études d'association et les études de cartographie de QTL n'expliquent que peu de la variabilité phénotypique suggérant ainsi un déterminisme génétique complexe des voies de la biosynthèse des composants du bois.

Dans l'objectif de compléter les connaissances disponibles sur la variabilité génétique et le déterminisme génétique de la composition chimique du bois chez le peuplier et notamment de définir des gènes candidats positionnels pour le contrôle de ces caractères, nous avons entrepris

une détection de QTL dans un pédigrée de cartographie génétique de peuplier noir. En outre, pour aller plus loin dans l'intérêt potentiel de ces caractères et de leurs déterminismes pour l'amélioration génétique, nous nous sommes aussi intéressés à la qualité technologique du bois pour une conversion en bioéthanol. Le phénotypage de ce pédigrée pour les différentes composantes de la qualité chimique et technologique du bois a été réalisé de façon indirecte par spectrométrie proche infrarouge (SPIR) après établissement d'équations de calibration sur un sous-échantillon. L'utilisation de ces équations de calibration dans le pédigrée nous a permis d'évaluer la variabilité génétique observée pour ces propriétés du bois et de cartographier des QTL pour ces caractères. Actuellement, aucune étude de cartographie de QTL pour la qualité technologique du bois chez le genre peuplier n'a été publiée.

C. Matériel et méthodes

1. Préparation des échantillons

Les échantillons de cette étude proviennent d'une famille de plein-frères constituée de 324 individus. Cette famille est issue de croisements contrôlés entre deux parents *P. nigra*. Le parent femelle 71072-501 échantillonné en 1972, vient d'une population de Savoie. Le parent mâle BDG vient de la vallée de la Garonne. La famille a été plantée dans la pépinière de l'INRA d'Orléans en 2004 avec une densité de 1,20m x 0,5m (assimilé à un dispositif en TtCR) à raison de 2 répétitions par clone. Depuis la plantation, les arbres sont recépés chaque année.

L'évaluation des propriétés du bois porte sur 479 échantillons qui correspondent à 271 clones fois 1,77 répétitions car certaines cépées étaient mortes lors de la récolte effectuée en mars 2010. Des sections de tige d'environ 50 cm ont été prélevées à la base de la tige dominante des cépées (en omettant les 20 premiers cm), puis séchées et intégralement broyées au moyen d'un broyeur à couteaux avec une grille de 1 mm. Le broyat ainsi obtenu a été tamisé afin d'enlever les fractions de poudre inférieure à 50 µm et supérieure à 1000 µm indésirables pour certains dosages chimiques. Une étude préliminaire sur un échantillon choisi au hasard montre que ces fractions sont négligeables car elles représentent moins de 1 % de la matière sèche.

2. Acquisition des spectres et prétraitements statistiques

Les échantillons broyés et tamisés (fraction 50 - 1000 µm) ont été analysés par spectrométrie proche infrarouge au moyen d'un spectromètre *Spectrum 400* de *Perkin Elmer* (*Massachusetts, USA*). Dix à 15 grammes de poudre par échantillon ont été déposés dans des coupelles en quartz. L'acquisition des spectres a été réalisée sur 64 points de l'échantillon et on obtient au final une moyenne des 64 spectres d'acquisition par échantillon. L'intervalle des

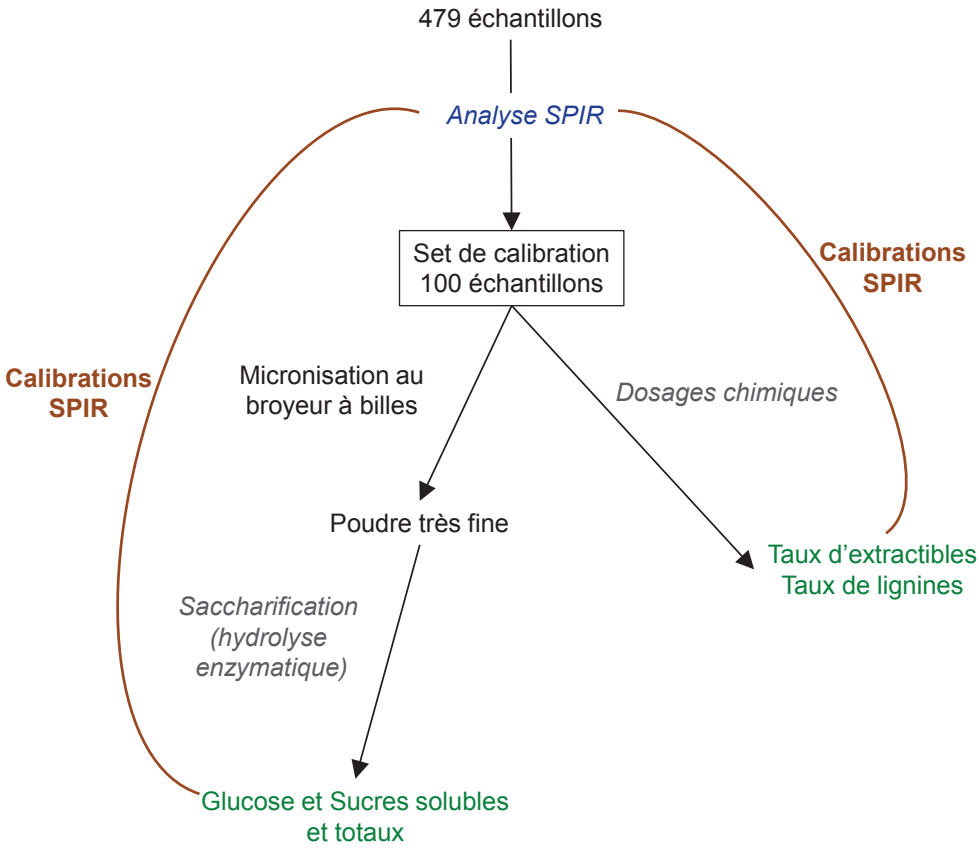


Figure V.2 : Représentation schématique des analyses (dosages et spectrométrie) effectuées pour phénotyper les caractères de qualité du bois des 479 échantillons du pédigrée F1 *P. nigra*.

spectres acquis est compris entre 4000 et 10000 cm^{-1} avec un pas de 2 cm^{-1}. Ainsi, 3001 valeurs d'absorbance sont obtenues pour chaque échantillon. Les spectres ont ensuite été découpés de façon à retenir les valeurs entre 4000 à 8000 cm^{-1}. Ils ont subi ensuite différents prétraitements statistiques afin d'améliorer la qualité du signal : normalisation (centrage et réduction), et dérivation première (fenêtre de 36 cm^{-1}), et seconde (fenêtre de 60 cm^{-1}), au moyen du filtre de Savitzky et Golay (Savitzky and Golay 1964), ainsi qu'une combinaison de ces prétraitements statistiques pour obtenir 8 modalités : spectres bruts, normalisation, dérivée première, dérivée seconde, normalisation puis dérivée première, normalisation puis dérivée seconde, dérivée première puis normalisation, et dérivée seconde puis normalisation. Toutes ces analyses ont été effectuées sous le logiciel R (R Development Core Team, 2012), notamment au moyen de la fonction *sgolayfilt* de la bibliothèque «*signal*» pour les dérivations des spectres. Les spectres acquis pour l'ensemble des échantillons et ayant subi une normalisation suivie d'une dérivée première ont été analysés avec l'algorithme de Kennard et Stone (Kennard and Stone 1969) de façon à définir un sous-ensemble de 100 échantillons qui soient représentatifs de la gamme de variabilité spectrale de l'ensemble des échantillons. Ces 100 échantillons constituent un «set» de référence sur lesquels les dosages chimiques ainsi que les calibrations SPIR ont été effectuées.

3. Analyses chimiques et potentiel de saccharification

Il était initialement prévu d'effectuer les analyses chimiques de référence dans un laboratoire privé extérieur avec la méthode dite Van Soest (Van Soest et al. 1991) utilisée pour l'analyse des herbacées. Des analyses préliminaires sur des poudres de bois de peupliers hybrides, effectuées par le laboratoire agronomique de Normandie (LANO) avec cette méthode n'ont pas donné satisfaction notamment pour les teneurs en lignines car les dosages n'étaient pas répétables. Ces essais ont montré que la méthode Van Soest ne semblait pas appropriée pour les dosages des propriétés chimiques du matériau bois. D'autres méthodes ont alors été utilisées pour effectuer des dosages plus précis des composants pariétaux. Ces dosages se déroulent en 2 étapes : (1) préparation du résidu pariétal et détermination de la teneur en extractibles, (2) dosages des teneurs en lignines solubles et insolubles «Klason». En omettant les cendres (éléments minéraux) qui représentent dans nos échantillons environ 1% de la matière, il est alors possible de déduire la teneur en holocellulose par différence : Holocellulose = 100 – (Extractibles + Lignines). Par ailleurs, afin d'évaluer la qualité technologique des échantillons, leur potentiel de saccharification a été évalué par dosage des sucres réducteurs et du glucose avant (solubles) et après hydrolyse enzymatique (totaux) (**Figure V.2**).

1 g poudre de bois

⇩ *2 bains de 30 min à l'éthanol à 96%*

Poudre de bois sans polyphénols

⇩ *2 bains de 30 min à l'eau*

Poudre de bois sans polyphénols et sans tannins = résidu pariétal

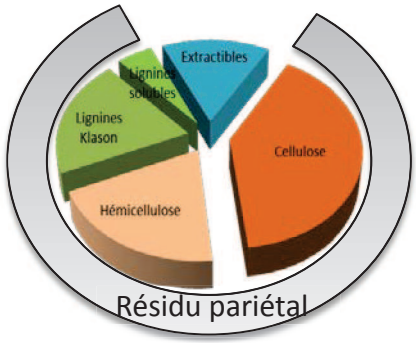


Figure V.3 : Représentation schématique du protocole de préparation du résidu pariétal permettant le dosage de la teneur en extractibles

300 mg poudre de bois extraite sèche

⇩

Hydrolyse à l'acide sulfurique à 72%
Bain marie 20°C - 2 heures

⇩

Hydrolyse à l'acide sulfurique à 3%
Autoclavage à 120°C - 1 heure

⇩

Filtration et rinçages

⇙ ⇘

Lignines Klason insolubles (K)

⇩

Étuve 60°C 12h puis étuve 103°C 2h
Dessiccateur

⇩

Pesée (K)

⇘

Lignines solubles dans filtrat (S)

⇩

Dilution au 1/20

⇩

Spectromètre UV
Absorbance 205 nm (S)

⇙

(K) + (S) = Lignines

Figure V.4 : Représentation schématique du protocole de dosage de la teneur en lignines

a) *Teneur en extractibles*

La méthode utilisée dérive de la méthode traditionnelle d'extraction à l'aide d'un appareillage *Soxhlet* (**Figure V.3**). Le point le plus important de modification de notre méthode basée sur la gravimétrie est le nombre d'échantillons extraits à la fois dans un même bain et l'usage de sachets spéciaux permettant de peser directement la poudre de bois extraite après séchage des sachets.

La préparation du résidu pariétal permet d'extraire les composés solubles de la poudre de bois qui interférent au moment du dosage des lignines et de la cellulose (Thammasouk et al. 1997). On peut ainsi évaluer le taux d'extractibles présents dans la poudre de bois ainsi que la fraction de résidus pariétaux sur lesquels seront dosées les lignines. Une fois les échantillons broyés et tamisés, on pèse un gramme de poudre de bois dans des sachets de fibre de verre thermo-soudables (*Ankom Technology, Macedo, USA*) de porosité 25 µm. Six à huit sachets sont introduits dans un ballon de 2 L placé sur un chauffe ballon. On réalise successivement deux bains de 500 mL d'éthanol à 96% porté à ébullition pendant 30 minutes chacun. Les polyphénols, de type flavonoïdes et acides phénols, étant solubles dans l'alcool, sont ainsi éliminés. Puis, après récupération des sachets, on réalise deux autres bains successifs de 500 mL d'eau ultra-pure portée à ébullition pendant 30 minutes chacun. On élimine ainsi les tanins, composés phénoliques solubles dans l'eau. Les sachets sont ensuite placés dans un four à 65°C pendant une nuit, puis à 105°C pendant 2 heures et ils sont pesés pour calculer la perte de masse et donc le taux d'extractibles contenus dans le bois.

b) *Dosage des lignines*

La méthode utilisée pour analyser les lignines dérive de la méthode dite «Uppsala Method» (Theander et al. 1995) (**Figure V.4**). On pèse 300 mg de résidu pariétal dans des bouteilles "*Schott",* dans lesquelles on ajoute 5 mL d'acide sulfurique à 72% sous hotte, puis on pratique une agitation mécanique très douce pendant 2 heures dans une salle tempérée à 20°C. Cette attaque acide permet de détruire les parois cellulaires du bois. On ajoute précisément 193,5 g d'eau ultra-pure pour diluer le mélange à 3%. Les bouteilles sont placées ensuite dans un autoclave à 120°C pendant 1 heure. Cette étape permet de finaliser la destruction des parois cellulaires et solubiliser la cellulose et l'hémicellulose. Après refroidissement des bouteilles, on filtre le mélange sous vide sur une rampe de filtration et dans des creusets de porosité allant de 16 à 40 µm. On récupère le filtrat qui est ensuite dilué au 1/15 pour mesurer l'absorbance par spectrophotométrie UV en cuve de quartz à 205 nm (longueur d'onde d'absorption de la lignine), ainsi on détermine la lignine soluble en utilisant une constante d'absorptivité de 110 $M^{-1}.cm^{-1}$

(TAPPI 1991).

Pour terminer la filtration, on fait passer environ 500mL d'eau ultra-pure sur chaque creuset pour éliminer toute trace de lignine soluble. Les creusets sont ensuite mis au four pendant une nuit à 65°C puis 2 heures à 105°C pour sécher la poudre de bois qui sera pesée. On obtient ici la lignine insoluble, appelée « lignine Klason ». On obtient alors la teneur en lignines totales en sommant les teneurs en lignines solubles et Klason et en l'exprimant en fonction de la matière sèche.

c) *Potentiel de saccharification*

Les mêmes échantillons analysés pour leur teneur en extractibles et en lignines ont été micronisés (< 50 µm) au broyeur à billes et envoyés à la plateforme de l'UMR biotechnologie des champignons filamenteux de Marseille pour évaluer leur potentiel de saccharification.
Cette plateforme entièrement robotisée est dédiée au criblage de la biodiversité enzymatique, et permet d'effectuer de façon automatique des saccharifications sans prétraitement chimique de la matière (Navarro et al. 2010). Les échantillons ont été tout d'abord analysés pour doser d'une part les sucres réducteurs solubles, et d'autre part le glucose soluble. Ensuite, ces échantillons subissent une hydrolyse par un cocktail enzymatique industriel (32 enzymes différents) extrait et purifié du champignon ascomycète *Trichoderma reesei.* Les sucres réducteurs totaux et le glucose total libérés après cette hydrolyse enzymatique sont dosés en triplicat. Le dosage des sucres réducteurs (cellobiose, xylose, arabinose ou cellulose cristalline, etc…) est effectué avec le DNS (3,5-acide dinitrosalicylique). Le DNS en présence de sucres réducteurs développe un produit réduit de couleur rouge-brune, l'acide 3-amino-5 nitrosalicylique qu'on dose par colorimétrie. Le dosage du glucose est effectué avec la méthode enzymatique « GOD POD » GOD : Glucose Oxydase et POD : Péroxydase, le produit final du dosage est une « quinoeimine dye » qui absorbe la lumière à 505 nm. De nouvelles variables ont alors été calculées, afin d'obtenir les teneurs en sucres et en glucose hydrolysés par différence entre les sucres et le glucose dosés après (appelés totaux) et avant (appelés solubles) hydrolyse enzymatique. On a aussi calculé la proportion de glucose dans les sucres solubles, hydrolysés et totaux, afin d'évaluer de façon indirecte la part de sucres en C6 par rapport aux sucres totaux et donc la cellulose par rapport à l'holocellulose.

4. Modèles de calibration

Les modèles de calibration ont été élaborés au moyen du package «pls» du logiciel R (Mevik and Wehrens 2007), en utilisant la régression par les moindres carrés partiels (PLS, Partial Least Squares) sur les données spectrales prétraitées. Notre approche de construction du

meilleur modèle de prédiction se fait en trois étapes. Tout d'abord, on identifie les observations aberrantes 'outliers' qui sortent de la gamme de calibration. Ce sont les observations dont le résidu de validation croisée sort de la distribution théorique attendue. Ensuite, on sélectionne un sous-échantillon de nombres d'ondes « pertinents » en utilisant l'algorithme CARS (competitive adaptive reweighted sampling) (Li et al. 2009). Cette technique permet de rechercher un sous ensemble de nombres d'onde minimisant l'erreur de validation croisée. Enfin, à partir du jeu de données filtré pour les observations aberrantes et seulement avec les nombres d'onde précédemment sélectionnés, des modèles sont construits. Le nombre de composantes de ces modèles est déterminé selon le critère de Wold (R=1) par validation croisée répétée (Monte - Carlo) avec k = 4 et 1000 répétitions. En outre, au cours de la validation croisée, diverses statistiques sont estimées afin d'évaluer la qualité du modèle sélectionné, comme le coefficient de détermination de validation croisée (R^2cv), l'erreur moyenne de prédiction en validation croisée (RMSEcv) et le critère RPD (ratio entre écart-type de la variable de référence à calibrer et RMSEcv). Pour une variable à calibrer, parmi les modèles construits pour chaque prétraitement statistique, celui maximisant le R^2cv ou minimisant le RMSEPcv a été sélectionné comme modèle de calibration.

5. Set de validation

Après établissement des modèles et prédiction des variables sur l'ensemble des échantillons du pédigrée, 16 échantillons ont été choisis pour représenter la gamme de variabilité de la teneur en lignines et ont subi toutes les analyses chimiques précédemment effectuées sur le set de référence pour la construction des modèles de calibration. Ce set de validation sert à comparer les écarts entre valeurs expérimentales et entre valeurs prédites par les modèles de calibration.

6. Héritabilités et corrélation entre caractères

Une fois les modèles de calibration validés, leur utilisation sur l'ensemble des échantillons du pédigrée nous a permis de prédire les propriétés du bois de l'ensemble des échantillons. Afin d'estimer l'héritabilité au sens large des caractères, les données ont été analysées au moyen du modèle linéaire mixte suivant :

$P_{ij} = \mu + G_i + \varepsilon_{ij}$,

où P_{ij} est la valeur phénotypique du caractère prédit pour la répétition j du génotype i, μ est la moyenne des valeurs phénotypiques prédites sur l'ensemble du pédigrée, G_i est l'effet aléatoire du génotype i et ε_{ij} est l'effet aléatoire du micro-environnement sur la répétition j du génotype i.

Tableau V.3 : Analyse exploratoire de la variabilité phénotypique des caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra.*

	Moyenne	Écart-type	Minimum	Maximum
Dosages chimiques				
Extractibles (%)	8,88	2,74	3,19	15,71
Lignines (%)	25,11	1,60	20,78	28,98
Holocellulose (%)	66,00	2,49	60,47	72,35
Solubles (avant hydrolyse)				
Sucres [µmol/mg]	0,16	0,08	0,08	0,38
Glucose [µmol/mg]	0,07	0,02	0,05	0,14
Proportion du Glucose	0,54	0,14	0,28	0,82
Totaux (après hydrolyse)				
Sucres [µmol/mg]	0,50	0,08	0,38	0,75
Glucose [µmol/mg]	0,27	0,05	0,19	0,39
Proportion du Glucose	0,55	0,09	0,37	0,72
Hydrolysés (Totaux – Solubles)				
Sucres [µmol/mg]	0,35	0,06	0,24	0,51
Glucose [µmol/mg]	0,20	0,05	0,11	0,33
Proportion du Glucose	0,56	0,07	0,42	0,72

Préalablement, les données individuelles ont été transformées avec la procédure BoxCox pour assurer l'homoscédasticité et la normalité des résidus du modèle linéaire. Ces modèles ont été établis pour chaque caractère au moyen de la bibliothèque Asreml du logiciel R (Butler et al. 2009) qui permet d'obtenir les estimations des variances génotypiques (σ^2_G) et résiduelles (σ^2_ε) par maximum de vraisemblance restreint ainsi que la matrice de covariance des composantes de la variance. Les valeurs d'héritabilité au sens large ont alors été calculées de la façon suivante : $H^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_\varepsilon)$. L'erreur standard des valeurs d'héritabilité a été calculée en utilisant la méthode delta (Lynch and Walsh 1998)

Les corrélations entre les moyennes génotypiques ont été estimées par la méthode de Pearson.

7. Détection de QTL

La détection des QTL a été effectuée sur les moyennes clonales des données phénotypiques transformées et sur trois cartes génétiques; deux cartes génétiques jalons parentales et une carte consensus jalon. Il est ainsi possible de détecter les effets d'additivité femelle et mâle sur les cartes femelle et mâle respectivement, alors que sur la carte consensus, il est possible de détecter outre les effets d'additivité, les effets de dominance entre les classes alléliques.

Dans une première étape, les QTL putatifs ont été détectés en Interval Mapping (IM) grâce au logiciel MapQTL6.0 (Van Ooijen 2009). Pour identifier ces QTL, on effectue des détections de QTL sur des ré-échantillonnages aléatoires des données (1000 permutations) (Churchill and Doerge 1994) afin d'obtenir une distribution des LOD scores maximums obtenue pour chaque permutation. Cette distribution permet de déterminer une valeur du LOD score seuil correspondant à un risque global alpha d'erreur sur tout le génome égal à 5%. Les marqueurs flanquant le QTL ainsi détectés, sont considérés comme des cofacteurs et sont intégrés dans un modèle multiple MQM (Multiple QTL Mapping). La détection des QTL en MQM permet d'affiner les intervalles de confiance et de détecter d'autres QTL à faible effet. Dans une deuxième étape, une détection de QTL épistatiques a été effectuée sous R sur les moyennes clonales des données phénotypiques transformées. Pour cela, les combinaisons entre les différentes classes alléliques ne respectant pas une ségrégation 1:1:1:1 (p=5%) sont d'abord éliminées du jeu de données avant d'analyser toutes les paires de marqueurs restantes avec le modèle linéaire suivant :

$$Y_{ijk} = \mu + M_i + N_j + (MN)_{ij} + \varepsilon_{ijk},$$

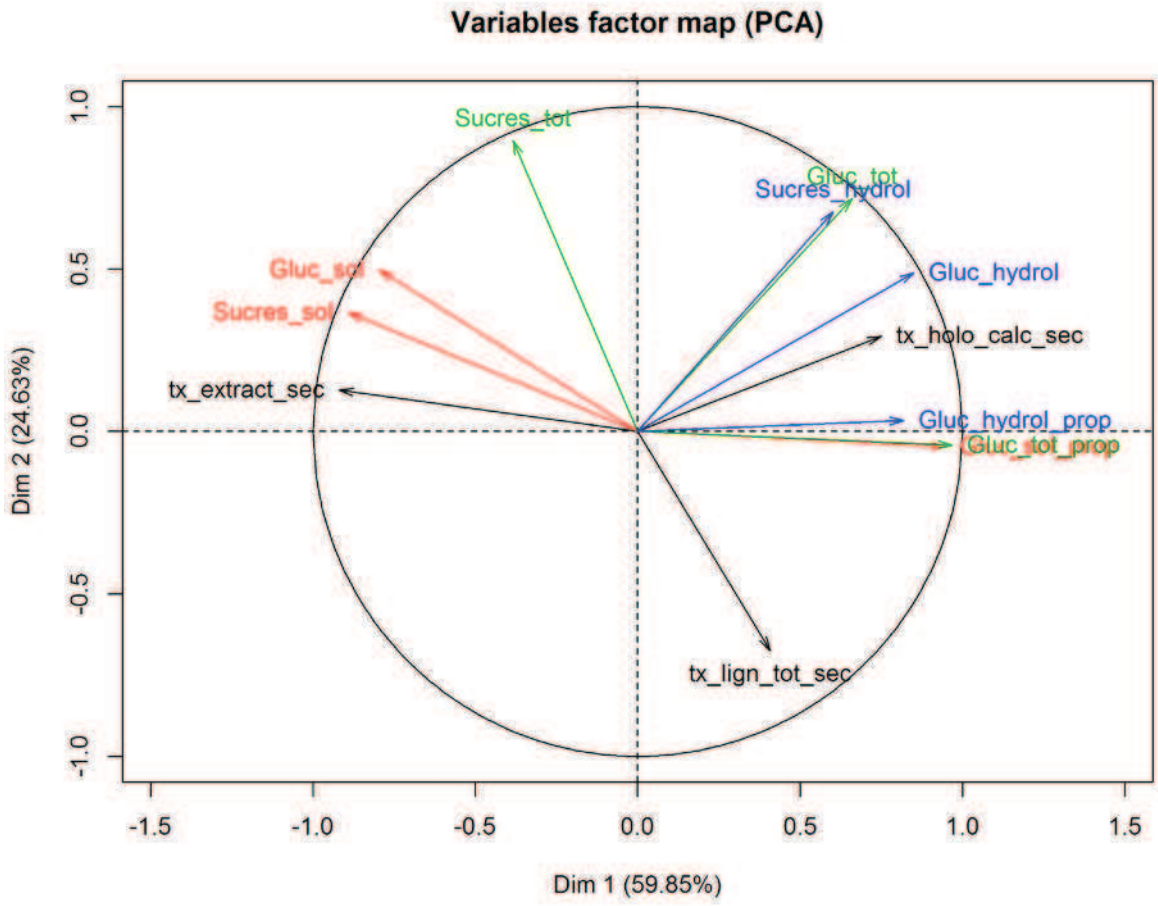


Figure V.5 : Analyse en composantes principales des caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra*. (tot : total, sol : soluble, hydrol : hydrolysé ; Gluc : glucose ; lign : lignines ; holo : holocellulose).

où Y_{ijk} est la valeur phénotypique du génotype k porteur de l'allèle i au marqueur M et de l'allèle j au marqueur N, μ est la moyenne, M_i and N_j sont les effets fixes des allèles i et j aux marqueurs M and N, $(MN)_{ij}$ est l'effet fixe d'interaction entre les allèles i et j des marqueurs M et N, et ε_{ijk} est la résiduelle.

On a effectué trois tests statistiques pour tester : 1/ les effets marginaux des marqueurs M et N ; 2/ l'effet d'interaction entre les marqueurs M et N ; 3/ les effets marginaux des marqueurs et leur interaction (modèle complet *vs.* modèle nul). Pour chaque test, un seuil alpha égal à 5% et corrigé avec la méthode de Bonferroni a été utilisé pour identifier les marqueurs dont l'effet était significatif.

Les pourcentages des variances expliquées par les QTL préalablement détectés sous MapQTL et R ont été évalués sur les moyennes clonales des données phénotypiques non transformées par le logiciel R en utilisant un modèle linéaire global, c'est-à-dire qui inclut tous les marqueurs précédemment détectés. Ces pourcentages sont en effet plus explicites sur les valeurs non transformées.

D. Résultats

1. Analyse exploratoire de la variabilité phénotypique dans les échantillons de référence

Avec des distributions normales selon le test de normalité de Shapiro-Wilk (alpha = 5%), les teneurs en extractibles, lignine et holocellulose représentent en moyenne 8,88%, 25,11% et 66% de la matière sèche totale respectivement. Le glucose est représenté en moyenne à plus de 50% dans toutes les fractions de sucres (solubles, hydrolysées et totaux) avec des échantillons présentant plus de 70% de glucose dans les sucres totaux et hydrolysés (**Tableau V.3**).

Les phénotypes des 100 échantillons désignés comme référence ont été analysés par une ACP. Trois groupes de variables se distinguent selon les deux premiers axes de l'ACP qui expliquent ensemble 86% de la variabilité observée : 1/ Les sucres solubles et les extractibles, 2/ Les sucres hydrolysés et l'holocellulose et 3/ la lignine et les sucres totaux (**Figure V.5**). Les extractibles sont corrélés positivement avec les sucres solubles, l'holocellulose est corrélée positivement avec la concentration en glucose hydrolysé ainsi que les proportions du glucose dans les sucres solubles, totaux et hydrolysés. La lignine est corrélée négativement avec les sucres totaux confirmant l'antagonisme qui existe entre la teneur en lignine et le potentiel de saccharification. Il existe aussi une corrélation positive entre la proportion de glucose dans les sucres solubles et la proportion de glucose dans les sucres totaux. Il n'y a pas de corrélation entre

Tableau V.4 : Modèles de calibration sélectionnés pour les caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra*.

Caractère	Prétraitement	Nb. VL	R^2_{train}	R^2_{cv} moy (sd)	$RMSE_{cv}$ moy (sd)	RPD_{cv} moy (sd)	Nb. "Outliers"	Nb. "Lambda"
Dosages chimiques								
Extractibles	Der 2nde	5	0.87	0.83 (0.01)	1.12 (0.03)	2.46 (0.07)	3	131
Lignines	Der 1ère	5	0.86	0.83 (0.01)	0.63 (0.02)	2.42 (0.06)	10	505
Holocellulose	Der 1ère	6	0.88	0.84 (0.01)	0.97 (0.03)	2.49 (0.08)	13	279
Solubles (avant hydrolyse)								
Sucres	Der 1ère	5	0.91	0.88 (0.01)	0.03 (0.001)	2.92 (0.09)	0	72
Glucose	Der 2nde & Norm	4	0.87	0.85 (0.01)	0.01 (0.0002)	2.60 (0.06)	0	46
Proportion du Glucose	Der 2nde & Norm	6	0.89	0.84 (0.01)	0.06 (0.002)	2.52 (0.10)	1	262
Totaux (après hydrolyse)								
Sucres	Der 2nde & Norm	5	0.83	0.78 (0.01)	0.03 (0.001)	2.17 (0.07)	20	251
Glucose	Der 1ère	6	0.86	0.80 (0.02)	0.02 (0.001)	2.27 (0.09)	9	105
Proportion du Glucose	Der 1ère	5	0.89	0.85 (0.01)	0.03 (0.001)	2.63 (0.09)	1	80
Hydrolysés (Totaux - Solubles)								
Sucres	Norm	6	0.82	0.77 (0.02)	0.02 (0.001)	2.10 (0.07)	27	262
Glucose	Der 1ère & Norm	6	0.87	0.83 (0.01)	0.02 (0.001)	2.41 (0.08)	9	67
Proportion du Glucose	Norm & Der 1ère	5	0.85	0.80 (0.01)	0.03 (0.001)	2.26 (0.07)	15	360

Prétraitement : prétraitement statistique des spectres (**Der** :dérivée, **Norm** :normalisation) ; **Nb. VL** : nombre de variables latentes dans le modèle PLS; **R^2_{train}** : R^2 d'apprentissage ; **R^2_{cv}** : R^2 de validation croisée avec son écart type entre parenthèse ; **$RMSE_{cv}$** : Erreur quadratique moyenne de validation croisée avec son écart-type entre parenthèses ; **RPD_{cv}** : Ratio de l'erreur de validation croisée par rapport à l'écart-type du caractère avec son écart-type entre parenthèses ; **Nb. "Outliers" :** nombre d'observations aberrantes ; **Nb. "Lambda" :** nombre de longueurs d'ondes sélectionnées par l'algorithme CARS et utilisées dans le modèle PLS.

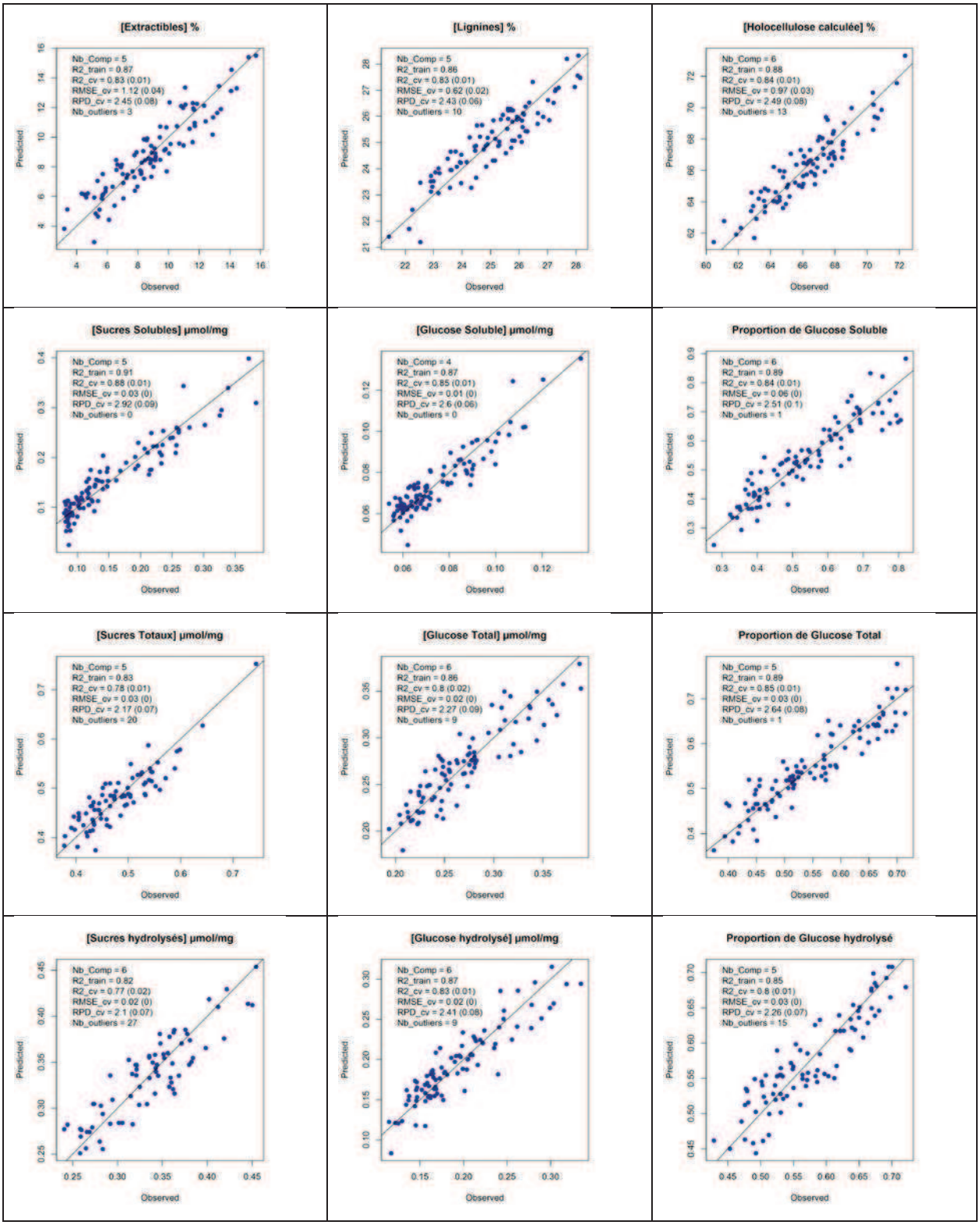


Figure V.6 : Représentation graphique des validations croisées de Monte-Carlo pour les modèles de calibrations sélectionnés pour les caractères mesurés sur les 100 échantillons de référence appartenant à un pédigrée F1 *P. nigra*.

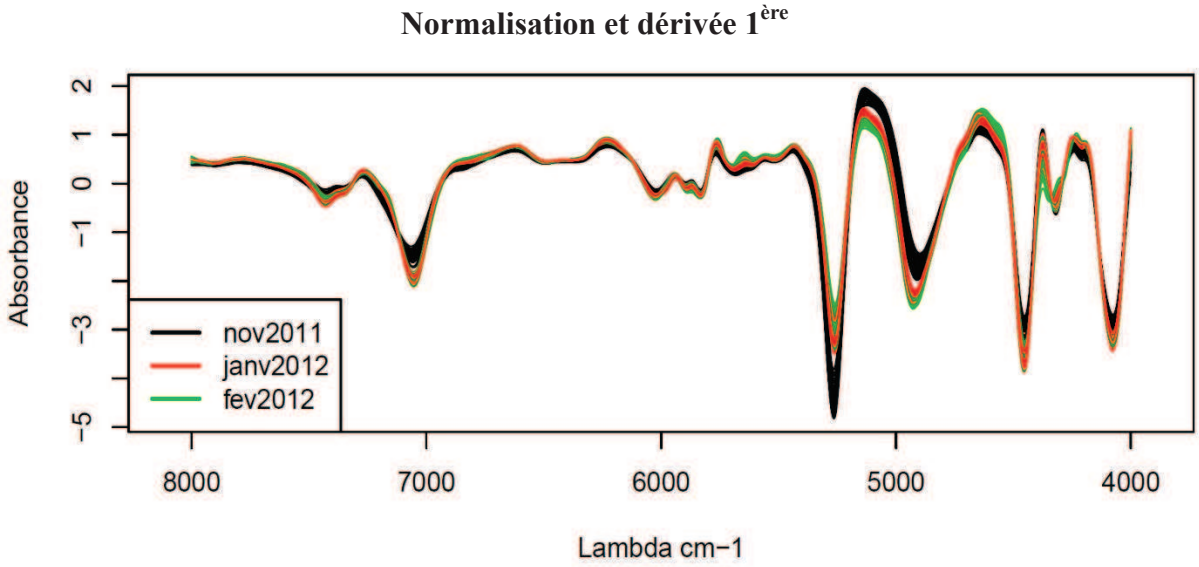


Figure V.7 : Illustration de l'effet « date » sur les spectres statistiquement pré-traités (normalisation et dérivée 1ère) acquis sur les 479 échantillons du pédigrée F1 *P. nigra*. Les 3 dates d'acquisition des spectres sont novembre 2011 (noir), janvier 2012 (rouge), et février 2012 (vert). Les 100 échantillons de référence ont tous été acquis en novembre 2011.

la lignine et l'holocellulose calculée. Ces résultats montrent la fiabilité des deux méthodes de phénotypage effectuées par deux techniques différentes et dans deux laboratoires différents.

2. Calibrations

Les modèles de calibration sélectionnés pour chaque caractère sont résumés dans le **tableau V.4** et la **figure V.6**. Ces modèles ont entre 5 et 6 variables latentes avec des R^2_{CV} de validation croisée allant de 0,77 pour les sucres hydrolysés à 0,85 pour le pourcentage de glucose dans les sucres totaux. La qualité de chaque modèle de calibration est aussi dépendante du nombre d'observations aberrantes, qui varie de 0 pour les sucres et le glucose solubles à 27 pour les sucres hydrolysés. Le RPD permet aussi d'évaluer la qualité des modèles. Il varie entre 2,10 pour les sucres hydrolysés et sucres solubles et 2,63 pour le pourcentage de glucose dans les sucres totaux. En comparant l'ensemble des modèles qui restent tous de bonne qualité, les meilleurs modèles réalisés correspondent à ceux obtenus pour les sucres solubles avec une seule donnée aberrante et des R^2_{CV} au-delà de 0,84. La qualité de ces modèles reflète la précision des mesures chimiques des sucres avant hydrolyse enzymatique. Après cette hydrolyse, les modèles construits sont de moins bonne qualité puisqu'on enlève plus de données aberrantes. Cet écart de précision peut s'expliquer par la nature biologique de l'hydrolyse enzymatique. Néanmoins, le modèle pour la proportion de glucose dans les sucres totaux reste très intéressant car avec seulement une donnée aberrante il se caractérise par un R^2_{CV} de validation croisée de 0,85 et un RPD de 2,63. Ce caractère est très intéressant parce qu'il représente la proportion de glucose dans le bois accessible par digestion enzymatique. C'est ce glucose qui serait disponible pour la fermentation en bioéthanol.

3. Phénotypes prédits, validations, héritabilités et corrélation entre caractères prédits

a) *Phénotypes prédits et validations*

L'analyse des spectres pris sur l'ensemble des échantillons a montré une structuration de quelques zones de spectres selon un effet correspondant aux 3 dates d'acquisition des spectres (**Figure V.7**). Cet effet était dû probablement à un taux d'humidité des échantillons différent aux trois dates voire au taux d'humidité de l'air ambiant lors de l'acquisition des données. Les spectres des 100 échantillons de référence avaient été tous collectés à la même date (novembre 2011), ce qui ne permettait pas de prendre en compte cet effet lors de la calibration. En revanche, il était possible d'ajuster les prédictions des modèles de calibrations à cet effet date, en prenant comme référence la date de prise des spectres des échantillons de référence. Pour illustrer cet effet ainsi que son ajustement sur les phénotypes prédits, l'intérêt d'un ajustement à l'effet date

Pas d'ajustement Ajustement à l'effet « date »

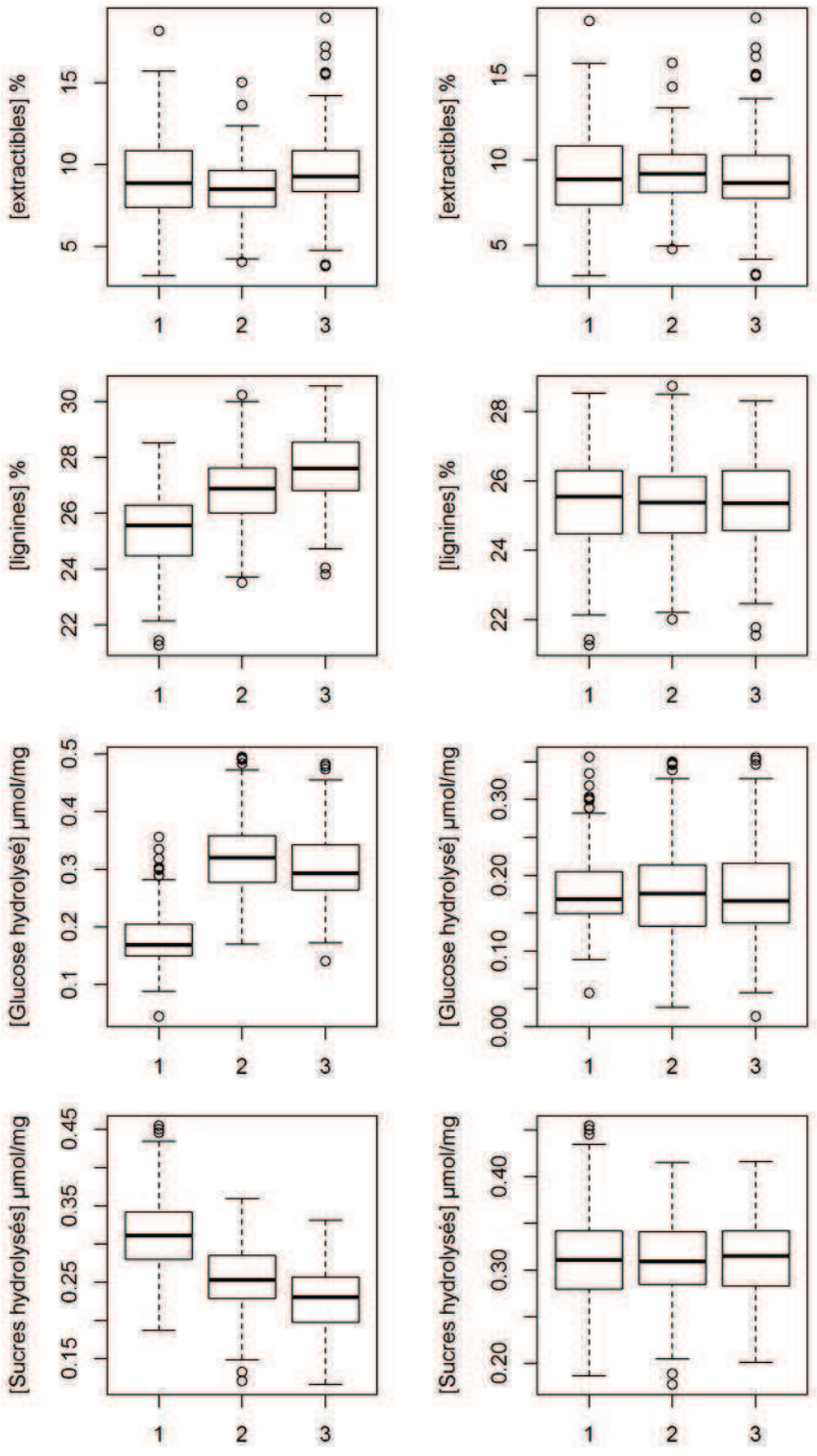


Figure V.8 : Représentation graphique de l'ajustement des valeurs prédites à l'effet « date » pour quelques caractères. Dates : 1 = Novembre 2011 ; 2 = Janvier 2012 ; 3 = Février 2012.

Tableau V.5 : Coefficients de détermination des régressions entre les valeurs mesurées et prédites (avec ou sans ajustement à l'effet « date » de prise des spectres) sur les 16 échantillons de validation du pédigrée F1 *P. nigra*.

Composé	R^2	
	Sans Ajustement	Avec Ajustement
Dosages chimiques		
Extractibles	0,80	0,74
Lignines	0,44	0,73
Holocellulose	0.39	0,82
Solubles (avant hydrolyse)		
Sucres	0,76	0,88
Glucose	0,76	0,81
Proportion du Glucose	0,53	0,68
Totaux (après hydrolyse)		
Sucres	0,27	0,31
Glucose	0,16	0,29
Proportion du Glucose	0,45	0,83
Hydrolysés		
Sucres	0,08	0,27
Glucose	0,08	0,15
Proportion du Glucose	0,07	0,47

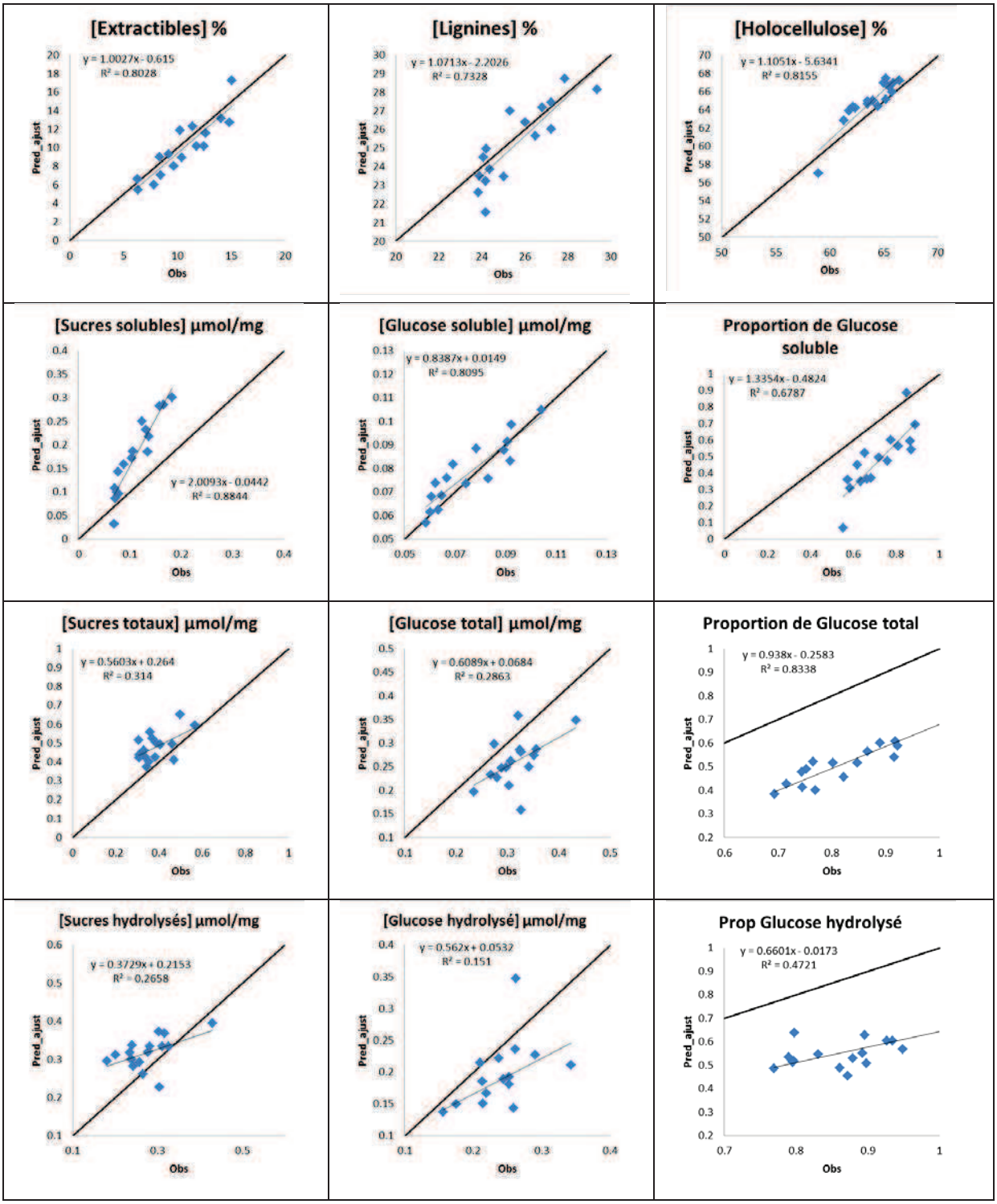


Figure V.9 : Représentation graphique des validations des modèles de calibrations sur 16 échantillons de validation du pédigrée F1 *P. nigra*. Les graphiques représentent les valeurs dosées (axe x) et les valeurs prédites ajustées à l'effet « date » (axe y). Le trait noir épais représente la droite y = x, tandis que le trait noir fin représente la droite de régression. Les équations de régression correspondantes ainsi que leurs coefficients de détermination sont indiqués.

est présenté pour 4 variables chimiques différentes dans la **figure V.8**. Si l'effet date semble plutôt limité pour le taux d'extractibles, il est significatif pour une ou plusieurs dates sur tous les autres caractères.

Afin de valider les modèles de calibration mais aussi l'ajustement à l'effet date des prédictions, des dosages ont été entrepris sur 16 échantillons supplémentaires. Ces échantillons ont été sélectionnés pour être représentatifs de la gamme de variation de la teneur en lignines prédite (après ajustement à l'effet date). Tout d'abord cette analyse nous a permis de valider l'ajustement des prédictions à l'effet date. En effet, hormis le taux d'extractibles, les corrélations entre valeurs prédites avec ajustement et valeurs observées étaient toujours meilleures que celles estimées entre valeurs prédites non ajustées et valeurs observées (**Tableau V.5**). Les régressions entre valeurs observées et prédites (avec ajustement) sont représentées de façon graphique en **figure V.9**. Ces validations étaient meilleures pour les teneurs en extractibles, lignines et holocellulose avec des R^2 de corrélation entre valeurs chimiques et valeurs prédites qui varient de 0,73 pour la lignine à 0,81 pour l'holocellulose. C'est le cas également pour les sucres solubles et la proportion de glucose total, même si dans la plupart des cas les prédictions dévient fortement de la droite y = x, impliquant dans certains cas des écarts entre les gammes de variation observées et prédites. Enfin pour les autres caractères, ces corrélations sont de moins bonne qualité même si dans certains cas (Glucose total, Sucres hydrolysés et proportion de glucose hydrolysé), on arrive à avoir de bons modèles. Cela est dû probablement à un faible nombre de points aberrants. Ces résultats sont donc assez conformes à ceux obtenus au cours des validations croisées.

b) *Héritabilités au sens large*

Le **tableau V.6** résume les moyennes et les écarts types observés pour les différents caractères prédits à l'aide des calibrations sélectionnées sur les 479 échantillons étudiés, sans transformation de variable. Les héritabilités au sens large ont été, quant à elles, calculées à partir de données transformées de façon à respecter les postulats du modèle linéaire. Les caractères les plus héritables sont les teneurs en holocellulose, en sucres totaux, en glucose total et en sucres hydrolysés avec plus de 60% de la variabilité phénotypique expliquée par la variance entre clones. Avec une gamme de variation s'étalant de 0,47 pour la proportion de glucose hydrolysé à 0,76 pour la teneur en sucres hydrolysés, toutes les héritabilités sont moyennes à élevées, montrant ainsi qu'il y a une grande part de la variabilité phénotypique due à la composante génétique. Ce résultat est encourageant pour la détection de QTL.

Tableau V.6 : Variabilité phénotypique et héritabilités au sens large de différentes propriétés chimiques du bois dans un pédigrée F1 de *P. nigra*. Les moyennes phénotypiques et leurs écart-types sont calculés sur les données prédites des 479 échantillons de l'étude. Les héritabilités au sens large (H^2) et leurs écart-type sont calculés à partir des données prédites et transformées par Box-Cox pour respecter les postulats du modèle linéaire.

	Moyenne	Écart type	H^2	(SE)
Dosages chimiques				
Extractibles	9,18	2,33	0,56	(0,05)
Lignines	25,35	1,35	0,56	(0,05)
Holocellulose	65,90	2,44	0,67	(0,04)
Solubles (avant hydrolyse)				
Sucres	0,16	0,06	0,58	(0,05)
Glucose	0,08	0,01	0,54	(0,05)
Proportion du Glucose	0,49	0,13	0,56	(0,05)
Totaux (après hydrolyse)				
Sucres	0,47	0,06	0,60	(0,04)
Glucose	0,25	0,05	0,67	(0,04)
Proportion du Glucose	0,53	0,08	0,56	(0,05)
Hydrolysés (Totaux – Solubles)				
Sucres	0,31	0,05	0,76	(0,03)
Glucose	0,18	0,06	0,54	(0,05)
Proportion du Glucose	0,56	0,06	0,47	(0,05)

Tableau V.7 : Matrice de corrélations phénotypiques entre caractères mesurés sur les 100 échantillons de référence (en-dessous de la diagonale) et entre les moyennes clonales des caractères prédits (au-dessus de la diagonale). Le niveau de significativité des corrélations est indiqué de la façon suivante : ns p-valeur > 0.05 ; * 0.05 ≤ p-valeur < 0.01 ; ** 0.01 ≤ p-valeur < 0.001 ; *** p-valeur ≤ 0.001. En gras et souligné |r| ≥ 0,9, en gras, 0,8 ≤ |r| < 0,9 et souligné 0,7 ≤ |r| < 0,8.

	Dosages chimiques			**Solubles (avant hydrolyse)**			**Totaux (après hydrolyse)**			**Hydrolysés (Totaux – Solubles)**		
Dosages chimiques	Extractibles	Lignines	Holocellulose	Sucres	Glucose	Proportion du Glucose	Sucres	Glucose	Proportion du Glucose	Sucres	Glucose	Proportion du Glucose
Extractibles	1	-0,42 ***	-0,78 ***	**0,89** ***	0,78 ***	**-0,90** ***	0,47 ***	-0,42 ***	**-0,81** ***	-0,41 ***	-0,62 ***	-0,60 ***
Lignines	-0,44 ***	1	-0,05 ns	-0,55 ***	-0,57 ***	0,34 ***	-0,78 ***	-0,11 *	0,37 ***	-0,29 ***	0,09 *	0,26 ***
Holocellulose	**-0,82** ***	-0,16 ns	1	-0,69 ***	-0,49 ***	0,74 ***	-0,08 ns	0,62 ***	0,75 ***	0,60 ***	0,72 ***	0,62 ***
Solubles (avant hydrolyse)												
Sucres	**0,85** ***	-0,58 ***	-0,56 ***	1	**0,89** ***	**-0,84** ***	0,61 ***	-0,35 ***	**-0,83** ***	-0,29 ***	-0,56 ***	-0,60 ***
Glucose	0,78 ***	-0,62 ***	-0,46 ***	**0,95** ***	1	-0,64 ***	0,78 ***	-0,01 ns	-0,57 ***	-0,08 ns	-0,24 ***	-0,31 ***
Proportion du Glucose	**-0,86** ***	0,43 ***	0,68 ***	**-0,88** ***	-0,73 ***	1	-0,28 ***	0,59 ***	**0,89** ***	0,50 ***	0,76 ***	0,70 ***
Totaux (après hydrolyse)												
Sucres	0,46 ***	-0,63 ***	-0,09 ns	0,67 ***	0,75 ***	-0,41 ***	1	0,38 ***	-0,23 ***	0,36 ***	0,14 **	-0,02 ns
Glucose	-0,46 ***	-0,12 ns	0,59 ***	-0,30 **	-0,14 ns	0,59 ***	0,41 ***	1	0,73 ***	0,76 ***	**0,88** ***	**0,81** ***
Proportion du Glucose	**-0,85** ***	0,41 ***	0,67 ***	**-0,84** ***	-0,74 ***	**0,92** ***	-0,43 ***	0,64 ***	1	0,55 ***	**0,85** ***	**0,89** ***
Hydrolysés (Totaux – Solubles)												
Sucres	-0,46 ***	-0,09 ns	0,56 ***	-0,37 ***	-0,22 *	0,56 ***	0,44 ***	**0,88** ***	0,48 ***	1	0,64 ***	0,65 ***
Glucose	-0,67 ***	0,09 ns	0,68 ***	-0,58 ***	-0,45 ***	0,77 ***	0,13 ns	**0,95** ***	**0,82** ***	**0,86** ***	1	**0,83** ***
Proportion du Glucose	-0,67 ***	0,28 **	0,55 ***	-0,6 ***	-0,56 ***	0,71 ***	-0,33 ***	0,63 ***	**0,91** ***	0,32 **	0,75 ***	1

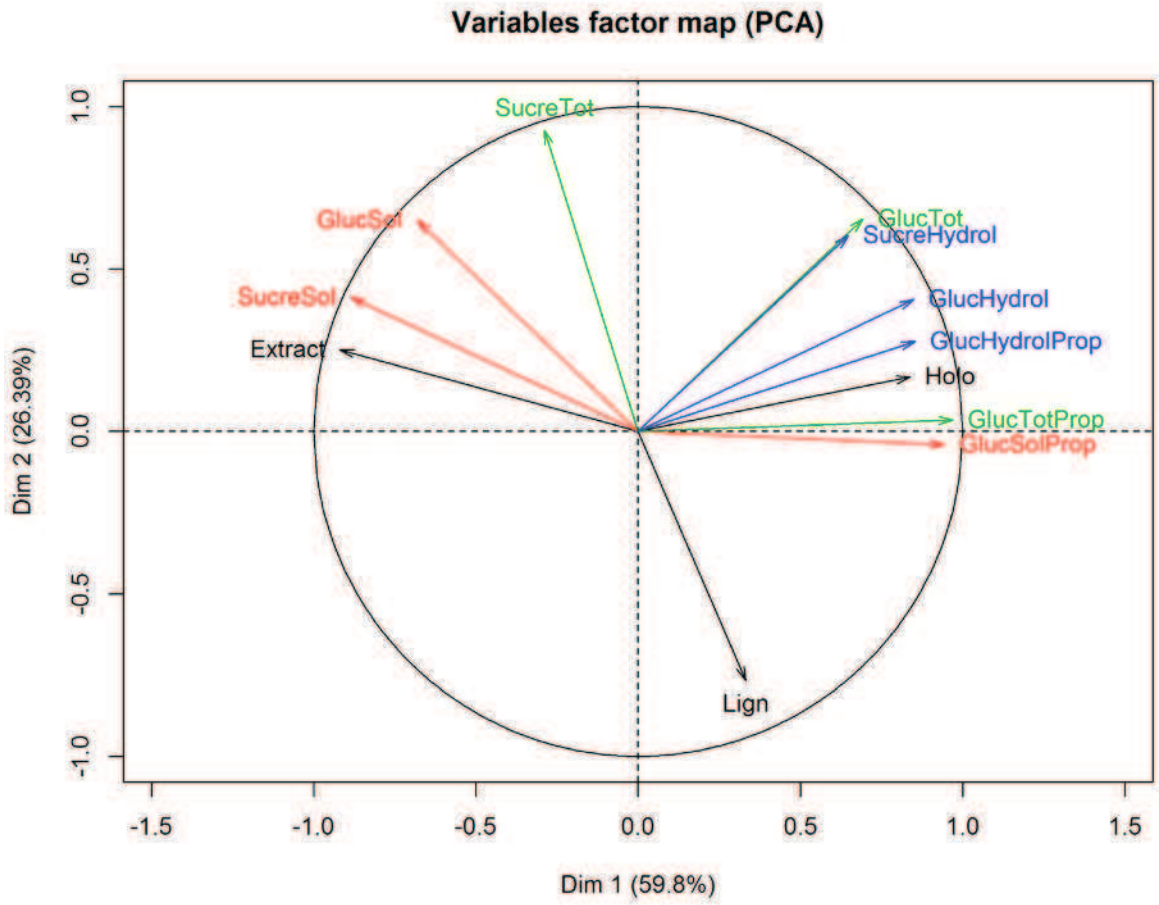


Figure V.10 : Analyse en composantes principales des moyennes clonales des caractères prédits sur les 479 échantillons du pédigrée F1 *P. nigra*. (tot : total, sol : soluble, hydrol : hydrolysé ; Gluc : glucose ; lign : lignines ; holo : holocellulose).

Tableau V.8 : Positions et statistiques associées aux QTL détectés en MQM (Multiple QTL Mapping) avec MapQTL sur les cartes jalon femelle, mâle et consensus d'un pédigrée F1 *P. nigra.*

Carte	Trait	GL	Position (cM)	IC (cM)	LOD_{max}	R^2	Marqueur associé
Femelle	Lignine	XII	86,2	56,2 - 95	3,08	4.8	LG_XII_16_194
		II	34	30,9 - 57,4	2,81	4,4	WPMS_11
	Sucres totaux	XII	83,2	56,2 - 94,8	3,01	4,8	LG_XII_16_194
Mâle	Sucres solubles	I	37,1	32,4 - 40,1	3,52	5,8	GCPM_877-1a
	Proportion du glucose dans les sucres totaux	I	35,1	32,4 - 40,1	2,83	4,7	GCPM_877-1a
Consensus	Lignine	II	68	64 - 86,5	4,17	6,8	WPMS_11
	Holocellulose	I	164,2	159,2 - 166,8	4,21	6,9	GCPM_877-1a
	Sucres solubles	I	163,2	159,2 - 166,8	4,93	8,0	GCPM_877-1a
	Sucres totaux	II	69,5	64 - 80,5	4,94	8,0	WPMS_11
	Proportion du glucose dans les sucres totaux	I	163,2	159,2 - 166,8	4,56	7,5	GCPM_877-1a
	Proportion du glucose dans les sucres hydrolysés	XIII	109,5	105,5 - 109,5	4,83	7,9	ORPM_297

GL : Groupe de liaison, **Position (cM)** : correspond au pic du LOD score ($\mathbf{LOD_{max}}$), **IC (cM)** : Intervalle de confiance du QTL correspondant au $LOD_{max} - 2$, $\mathbf{R^2}$: Pourcentage de la variation phénotypique expliquée par le QTL et calculé à partir des moyennes clonales des caractères prédits et transformés (Box-Cox). **Marqueur associé** : Le marqueur le plus proche du LOD_{max}.

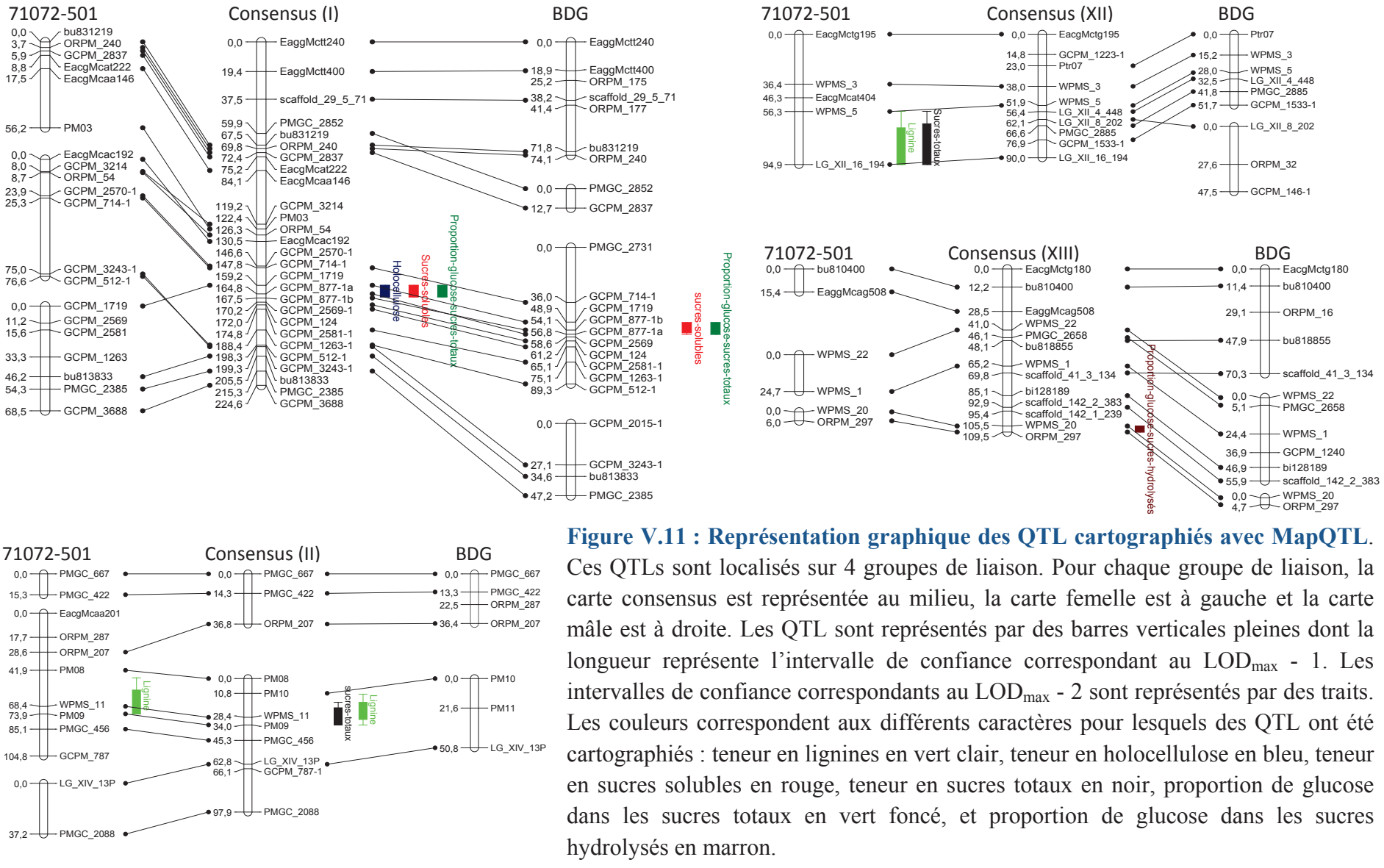


Figure V.11 : Représentation graphique des QTL cartographiés avec MapQTL. Ces QTLs sont localisés sur 4 groupes de liaison. Pour chaque groupe de liaison, la carte consensus est représentée au milieu, la carte femelle est à gauche et la carte mâle est à droite. Les QTL sont représentés par des barres verticales pleines dont la longueur représente l'intervalle de confiance correspondant au LOD_{max} - 1. Les intervalles de confiance correspondants au LOD_{max} - 2 sont représentés par des traits. Les couleurs correspondent aux différents caractères pour lesquels des QTL ont été cartographiés : teneur en lignines en vert clair, teneur en holocellulose en bleu, teneur en sucres solubles en rouge, teneur en sucres totaux en noir, proportion de glucose dans les sucres totaux en vert foncé, et proportion de glucose dans les sucres hydrolysés en marron.

c) *Corrélations entre caractères prédits*

Les corrélations phénotypiques qui existaient entres les caractères sur l'échantillon de 100 individus de référence utilisés pour la calibration sont conservées en signe et en amplitude après prédiction et au niveau des moyennes clonales (**Tableau V.7; Figure V.10**).

Il existe des corrélations entre caractères dues à leur dépendance dans les calculs. Mais il existe d'autres corrélations très intéressantes entre caractères estimés de façon indépendante. Ainsi comme précédemment sur les échantillons de référence, trois groupes de variables peuvent être identifiés (**Tableau V.7; Figure V.10**):

- Le taux d'extractibles est corrélé négativement avec la proportion du glucose dans les sucres solubles et dans les sucres totaux. Ce taux d'extractibles est corrélé positivement avec la concentration en sucres solubles. La teneur en sucres solubles est négativement corrélée avec la proportion de glucose dans les sucres totaux. La proportion de glucose dans les sucres solubles est corrélée positivement avec la proportion de glucose dans les sucres totaux. Cette dernière corrélation est très importante par ce qu'elle permettrait de prédire le taux de glucose libéré par l'hydrolyse enzymatique avec la simple mesure des teneurs en glucose et en sucres solubles.
- La teneur en holocellulose est corrélée positivement avec la proportion du glucose dans les sucres solubles, dans les sucres totaux et avec le taux de glucose hydrolysé. Elle est corrélée négativement avec la teneur en sucres solubles.
- Le taux de lignine est corrélé négativement avec la teneur en sucres totaux, confirmant l'antagonisme qui existe entre la teneur en lignine et le potentiel de saccharification. Il n'existe pas de corrélation au niveau des moyennes génotypiques entre teneur en lignine et en holocellulose malgré le fait que ces deux variables soient dépendantes par le calcul, ce qui prédit un déterminisme génétique indépendant.

4. Détection des QTL

La détection des QTL a été réalisée à l'aide des moyennes clonales calculées à partir des données individuelles prédites à l'aide des calibrations SPIR puis transformées (Box-Cox) pour répondre aux postulats du modèle linéaire. Elle a été effectuée sur des cartes parentales jalons séparées et sur la carte jalon consensus présentée dans le chapitre précédent.

Les QTL identifiés par MapQTL en analyse MQM sont présentés dans le **tableau V.8** et la **figure V.11**. Au total, 11 QTL ont été détectés pour 6 caractères différents. Deux QTL ont été identifiés uniquement sur la carte femelle, il s'agit des QTL pour les teneurs en lignines et en sucres totaux détectés sur le groupe de liaison XII. Trois QTL ont été détectés uniquement sur la

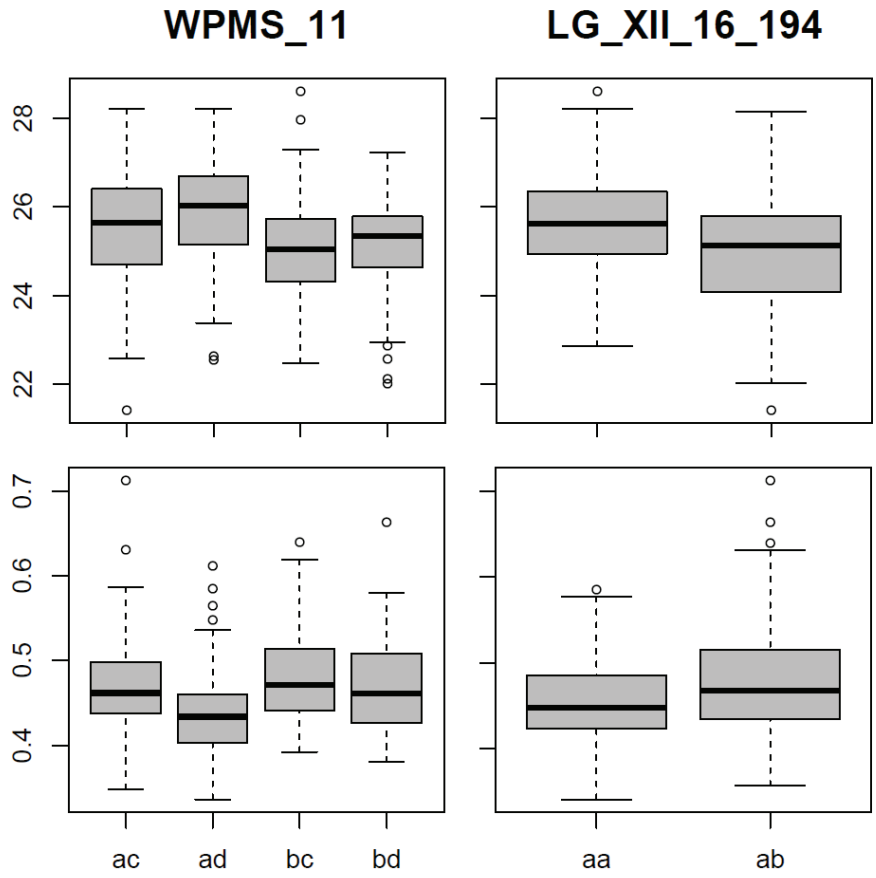


Figure V.12 : Représentation graphique des effets opposés des 2 QTL associés à la teneur en lignines et à la quantité de sucres totaux sous forme de boites à moustache pour les classes alléliques des 2 marqueurs les plus proches de ces QTLs, le marqueur SSR WPMS_11 du groupe de liaison II et le marqueur SNP LG_XII_16_194 du groupe de liaison XII.

carte consensus, il s'agit des QTL pour la teneur en holocellulose sur le groupe de liaison I, pour la teneur en sucres totaux sur le groupe de liaison II, et pour la proportion de glucose dans les sucres hydrolysés sur le groupe de liaison XIII. Les QTL restants ont été cartographiés simultanément sur la carte consensus et sur l'une des cartes parentales, il s'agit des QTL associés à la teneur en sucres solubles et à la proportion de glucose dans les sucres totaux sur le groupe de liaison I, et ceux associés à la teneur en lignine détectés sur le groupe de liaison II (**Figure V.11**). Les QTL pour la teneur en lignine et en sucres totaux qui co-localisent sur les groupes de liaison II et XII sont caractérisés par des effets opposés, ce qui est en accord avec la corrélation négative observée entre ces deux caractères (**Figure V.12**).

L'analyse sous R a permis de détecter 4 QTL supplémentaires dont 3 sont impliqués dans des interactions épistatiques (**Tableau V.9**). Pour la teneur en extractibles, un QTL situé sur le groupe de liaison IV et un marqueur non lié qui sont significatifs à la fois marginalement et en interaction. Pour la teneur en holocellulose, un QTL situé sur le groupe de liaison IV qui n'a pas d'effet marginal mais devient significatif en interaction avec un QTL situé sur le groupe de liaison I et qui a été précédemment identifié sous MapQTL. Pour la teneur en sucres totaux, un QTL situé sur le groupe de liaison IX.

Le pourcentage de variation phénotypique expliquée individuellement par les différents QTLs détectés avec MapQTL et R mais calculé à partir des moyennes génotypiques non transformées varie entre 2 % pour la teneur en sucres totaux et 16% pour la proportion de glucose dans les sucres hydrolysés (**Tableau V.9**). Pour les caractères pour lesquels plusieurs QTL ont été identifiés, nous avons aussi estimé les pourcentages de variation phénotypique totaux (**Tableau V.9**). Ainsi, les pourcentages de variation expliqués par caractère varient entre 9 % pour la teneur en sucres solubles et 16 % pour la teneur en holocellulose et la proportion de glucose dans les sucres hydrolysés (**Tableau V.9**).

E. Discussion

La technique de phénotypage indirecte à haut débit par SPIR, a permis d'analyser la variabilité génétique pour les propriétés chimiques et technologiques du bois de peuplier noir en vue de sa conversion en bioéthanol. On a montré que ces caractères sont sous contrôle génétique (héritabilités modérées à élevées) ce qui permet d'identifier des régions génomiques qui contrôlent la variabilité de ces caractères et d'ouvrir des perspectives en vue d'optimiser la qualité de la ressource par amélioration génétique.

Tableau V.9 : Ensemble des QTLs associés aux caractères de propriétés du bois détectés avec MapQTL et R dans un pédigrée F1 *P. nigra* pour leurs effets propres ou épistatiques.

Caractère	Marqueur(s)	GL	Détection	Origine	p-value individuelle	R2 individuel	p-value globale	R2 global
Dosages Chimiques								
Extractibles	PMGC_2270	IV	R	F	2,09E-03	3,6	9,18E-04	13,2
	EaggMcag61	NA	R	M	9,96E-03	2,7	1,20E-04	
	PMGC_2270 * EaggMcag61	IV * NA	R	F * M	NA	NA	2,94E-03	
Lignine	WPMS_11	II	MapQTL (F, C) & R	F & M	1,66E-04	7,4	1,45E-03	11,3
	LG_XII_16_194	XII	MapQTL (F) & R	F	4,49E-05	6,6	2,29E-04	
Holocellulose	GCPM_1719[a]	I	MapQTL (C) & R	M	9,09E-04	6,5	2,83E-04	15,6
	GCPM_1719[a]*ORPM_127	I*IV	R	F * M	NA	NA	1,20E-04	
Solubles (avant hydrolyse enzymatique)								
Sucres	GCPM_877-1a	I	MapQTL (M, C)	M	1,63E-05	9,4	1,63E-05	9,4
Totaux (après hydrolyse enzymatique)								
Sucres	WPMS_11	II	MapQTL (C) & R	F & M	1,90E-04	7,3	2,93E-04	14,7
	LG_IX_20_46	IX	R	M	1,58E-02	2,2	1,55E-03	
	LG_XII_16_194	XII	MapQTL (F)	F	4,06E-04	5,0	1,63E-03	
Proportion de glucose	GCPM_877-1a	I	MapQTL (M, C)	M	1,07E-05	9,8	1,07E-05	9,8
Hydrolysés (Totaux – Solubles)								
Proportion du glucose	ORPM_297	XIII	MapQTL (C)	F & M	2,36E-04	15,9	2,36E-04	15,9

[a] choisi à la place de GCPM_877-1a car il a 4 classes alléliques et son interaction avec ORPM_127 implique des effets femelles et mâles.
Marqueur : marqueur sélectionné pour représenter le QTL ; **GL** : groupe de liaison ; **Détection** : logiciel R ou logiciel MapQTL ; **Effet :** F (femelle), M (Mâle) ; **p-value individuelle** : p-valeur de l'effet du marqueur seul sur la moyenne clonale du caractère prédit et non transformé ; **R2 individuel** : pourcentage de la variation expliquée par le marqueur seul et calculé à partir de la moyenne clonale du caractère prédit et non transformé ; **p-value globale** : p-valeur partielle de l'effet du marqueur sur la moyenne clonale du caractère prédit et non transformé dans un modèle global qui réunit tous les QTL détectés pour un caractère ; **R2 global** : pourcentage de la variation expliquée par le modèle global et calculé à partir de la moyenne clonale du caractère prédit et non transformé.

1. Intérêt de la SPIR comme méthode de phénotypage indirecte des propriétés chimiques et technologiques du bois

La méthode de phénotypage indirecte à haut débit par SPIR va permettre aux généticiens de cribler un nombre important de génotypes pour en sélectionner les plus intéressants. Outre la validation croisée, la validation des modèles de prédiction par l'analyse d'échantillons complémentaires à l'aide des méthodes de référence a été effectuée. Les modèles à meilleur pouvoir de prédiction concernaient les teneurs en lignine, extractibles, et holocellulose, ainsi que les teneurs en sucres et glucose solubles avant hydrolyse enzymatique. Par contre, des modèles de moins bonne qualité ont été obtenus pour les teneurs en sucres et glucose après hydrolyse enzymatique ou issus de l'hydrolyse enzymatique. L'efficacité de chaque enzyme du cocktail dépend de plusieurs paramètres (pH, agitation, densité, température…) ce qui rend cette hydrolyse biologique plus variable qu'un dosage chimique. Ceci pourrait expliquer cette perte de qualité de prédiction.

L'utilisation de la SPIR comme méthode de phénotypage a été largement utilisée pour les cultures pérennes pour l'identification des QTL responsables de la qualité de la graine (Rami et al. 1998 ; Zhao et al. 2008; Gustafson and de Leon 2010) ou la qualité du fourrage (Cogan et al. 2005 ; Taleei et al. 2009). Chez les espèces forestières, les seules études de cartographie QTL de la qualité du bois basée sur une prédiction indirecte SPIR concernent l'eucalyptus (Thamarus et al. 2004 ; Thumma et al. 2010).

Avec 4 à 6 variables latentes en PLS et des R^2_{cv} qui varient entre 0,77 et 0,88, nos modèles de calibration sont de qualité comparable à celle des modèles présentés dans la littérature. Les modèles de prédiction des monomères de la lignine établis sur 267 hybrides transgéniques de peuplier ont 8 variables latentes en PLS, et des R^2_{cv} variant de 0,71 à 0,95 (Robinson and Mansfield 2009). De façon comparable, Zhou et al. (2011) ont établi des modèles de calibration en moyen infra rouge (SMIR) sur un pédigrée hybride interspécifique F2 de peuplier (Zhou et al. 2011). Leur modèle de prédiction de la teneur en lignine a 4 variables latentes en PLS et un R^2_{cv} de 0,67. Avec un R^2_{cv} de 0,81, leur modèle s'améliore en restreignant le nombre de longueur d'ondes spécifiques à la lignine, mais il a 12 variables latentes en PLS. Cependant, nos modèles restent de moins bonne qualité que ceux de Yeh et al. (2004) qui ont analysé des individus non apparentés du pin *taeda* pour prédire leur teneur en lignine. Leur modèle de calibration présente 5 variables latentes en PLS et un R^2_{cv} de 0,99 (Yeh et al. 2004). Hou et al. (2011) présentent aussi, sur des individus non apparentés de peuplier et eucalyptus, des modèles de calibration pour les teneurs en lignine, en holocellulose et en sucres issus d'hydrolyse enzymatique, et le ratio S/G. Ils retrouvent des R^2_{cv} qui varient entre 0,92 et 0,99 avec 7 à 8 variables latentes en

PLS (Hou and Li 2011). La différence entre nos modèles et ces deux modèles est due probablement à la gamme de la variabilité des individus à prédire qui est plus restreinte en intraspécifique (écart-type pour la teneur en lignine de 1,35 dans notre étude) par rapport à des individus appartenant à deux espèces (écart-type pour la teneur en lignines de 2,12 chez Hou et al. (2011)).

2. Influence de la composition chimique du bois sur le rendement en saccharification

L'ACP montre trois groupes distincts de descripteurs de la composition chimique du bois et de sa transformation par hydrolyse enzymatique : avec les sucres solubles, les extractibles d'un côté, les sucres totaux, les sucres hydrolysés et l'holocellulose de l'autre côté et un groupe avec deux composantes antagonistes, les sucres totaux et la lignine. Les mesures de ces caractères sont réalisées par des techniques différentes et dans deux laboratoires différents. Les résultats obtenus par les deux techniques différentes sont toutefois en parfaite cohérence avec les attendus biologiques de l'expérimentation. Les sucres solubles sont corrélés positivement avec les extractibles. Lors de la préparation des résidus pariétaux, les sucres solubles ainsi que les polyphénols et les tannins sont éliminés. Les sucres solubles correspondent donc à une fraction des extractibles.

Le dosage de la teneur en cellulose dans le bois est une méthode laborieuse. On a essayé de la mettre en place dans le laboratoire. Les mesures de la teneur en cellulose n'étaient pas répétables dû à l'hétérogénéité du matériau bois. En omettant les cendres, l'estimation du taux d'holocellulose, qui regroupe à la fois l'hémicellulose polymère de sucres C5 (majoritairement le xylose) et la cellulose polymère de sucres C6 (majoritairement le glucose), s'avère précise puisque ce taux est en corrélation positive avec les sucres hydrolysés et totaux.

La teneur en lignines qui regroupe les lignines solubles et la lignine « Klason » se retrouve à l'opposé des sucres totaux confirmant l'antagonisme entre sa teneur et celle des sucres (Studer et al. 2011) et la cohérence et fiabilité des mesures réalisées dans les deux laboratoires. Chez le peuplier, la teneur en lignine est corrélée négativement avec la teneur en sucres et en glucose totaux, tandis que cette relation négative n'a pas été détectée jusqu'ici chez le saule (Ray et al. 2012). En effet, dans cette dernière étude, il n'existe pas de corrélation entre les teneurs en lignine et en glucose libéré après un prétraitement acide, ni entre les teneurs en lignine et en glucose convertible en bioéthanol.

3. Le déterminisme génétique des propriétés chimiques et technologiques du bois est complexe

On a trouvé des valeurs d'héritabilité moyennes à élevées pour l'ensemble des caractères. Sur la même espèce mais dans un fond génétique plus hétérogène (17 familles de demi-frères issues de pollinisation libre), des valeurs comparables d'héritabilité au sens large ont été estimées pour les teneurs en lignines et en sucres (Guerra et al. 2013). Par contre, les valeurs d'héritabilité sont faibles dans une étude réalisée sur un pédigrée F1 d'un peuplier hybride (*P. trichocarpa x P. deltoides*) x *P. deltoides* (Novaes et al. 2009), ceci est dû probablement à l'âge des génotypes analysés dans cette étude. En effet, ils étaient récoltés après seulement 10 semaines de croissance en serre. Notre étude a été réalisée sur des sections de tige de 1 an qui n'étaient pas écorcées. Le taux d'écorce pourrait modifier essentiellement l'estimation du taux d'extractibles et conduirait à sous-estimer les teneurs en lignines et holocellulose évaluées à partir du résidu pariétal.

Les gammes de variabilités observées chez des espèces pures (*P. trichocarpa* et *P. nigra*) laissent espérer des gains génétiques significatifs en sélection phénotypique individuelle intra-famille voire en sélection récurrente sur plusieurs cycles de sélection. Les 5 meilleurs génotypes présentent un gain génétique de 7% par rapport à la moyenne pour la teneur en holocellulose, tandis que le gain génétique acquis par les 5 meilleurs génotypes au niveau de la teneur en lignine est une réduction de 11% par rapport à la moyenne.

Les héritabilités moyennes à élevées observées pour tous ces caractères montrent que la composante génétique dans la variation des caractères est importante. Toutefois, le petit nombre de QTL détectés et les faibles pourcentages de variation phénotypique expliquée (compris entre seulement 9 et 16 %) suggèrent un déterminisme génétique complexe de ces caractères impliquant probablement un grand nombre de loci à effets additifs faibles (modèle infinitésimal, Fisher 1918) et de possibles interactions épistatiques difficiles à détecter. En effet, dans notre situation peu favorable à la détection de ces interactions entre gènes, 3 cas d'interactions épistatiques détectées entre paires de marqueurs ont pu être mis en évidence. Des interactions plus complexes impliquant un plus grand nombre de gènes pourraient également expliquer une part significative de la variabilité génétique observée. Notre étude a permis de révéler un déterminisme génétique complexe de la plupart des caractères mesurés. Cette complexité est largement décrite dans la littérature (Zhong and Ye 2009 ; 2010 ; McCarthy et al. 2010 ; Du and Groover 2010 ; Drost et al. 2010 ; Dharmawardhana et al. 2010 ; Demura and Ye 2010).

Trois QTLs co-localisent sur le groupe de liaison I, ce sont des QTLs associés aux teneurs en holocellulose, en sucres solubles et à la proportion du glucose dans les sucres totaux. Ces

QTLs co-localisent aussi avec le QTL associé au ratio C6/C5 détecté précédemment chez le peuplier (Novaes et al. 2009). De même, le QTL détecté sur le groupe de liaison XIII et associé à la proportion de glucose dans les sucres hydrolysés co-localise avec les QTL associés aux ratio C6/C5, C6/lignine, à la teneur en lignine G et en lignine S détectés dans le même pédigrée de peuplier par Novaes et al. (2009). En revanche, aucun des QTL identifiés dans notre étude ne co-localisent avec les 4 QTL associés au taux de glucose hydrolysé chez le saule (Brereton et al. 2010). Ces co-localisations dans différents fonds génétiques sont très intéressantes puisqu'elles vont fournir de nouveaux candidats positionnels très intéressants pour de futures recherches de polymorphisme et études d'association.

4. Perspectives

Plusieurs étapes de validation de ces QTL sont maintenant à entreprendre :

(1) des validations liées à la représentativité inconnue d'un matériau bois jeune, récolté dans un seul environnement et des conditions de cultures particulières. Il existe deux plantations expérimentales âgées de 12 ans incluant une trentaine de génotypes de ce pédigrée qui pourraient offrir du matériel de validation lors de leur exploitation prochaine. Par ailleurs, une nouvelle récolte en parc à pied-mères de matériel plus âgé (2 voire 3 ans) pourrait également permettre d'évaluer la qualité d'une ressource représentative d'une récolte en TtCR et de comparer ses propriétés avec ou sans écorce.

(2) des validations portant sur le rendement en saccharification selon différentes méthodes de prétraitement. Si des teneurs élevées en holocellulose sont favorables à des rendements en saccharification élevés quel que soit la nature du prétraitement ; il convient de préciser l'influence du taux de lignine sur ces rendements dans différentes conditions de prétraitement. En effet, Studer et al. (2011) montre un effet négatif modéré de la teneur en lignine sur la libération des sucres totaux hydrolysés dans trois conditions différentes de prétraitements (180°C à 18 min, 160°C à 28 min, 140°C à 464 min). Le prétraitement industriel proposé par le projet FUTUROL qui finance ce travail est très puissant (steam explosion) et risque de niveler le niveau des sucres libérés en supprimant l'impact défavorable des teneurs élevées en lignine.

(3) des validations portant sur la stabilité de ces QTL dans différents fonds génétiques intra et interspécifiques. Les détections de QTL réalisées peuvent être rapidement étendues au plan de croisement factoriel 5x4 *P. nigra* qui inclut notre pédigrée F1. Pour chacun des deux parents mâle et femelle, ce sont respectivement 156 et 80 individus issus de combinaisons avec d'autres parents *P. nigra* qui sont d'ores et déjà disponibles en parc à pied-mères et plantations expérimentales. La validation en croisement interspécifique *P. deltoides x P. nigra* pourrait être

également réalisée à court terme pour les seuls QTL détectés sur le parent mâle BDG pour lequel, il existe 279 descendants issus de croisement avec 4 différentes femelles *P. deltoides*. Ces validations cumulant l'information sur un plus grand nombre d'évènements de recombinaison par parent permettraient également de réduire l'intervalle de confiance de ces QTL.

Pour étendre la liste des gènes candidats positionnels possibles tout en valorisant l'important polymorphisme présent naturellement chez l'espèce *P. nigra*, il est aussi prévu d'entreprendre une étude d'association pour s'approcher des régions génomiques candidates en exploitant le déséquilibre de liaison à courte distance entre candidat et marqueur. Pour définir les gènes sous-jacents aux QTL identifiés, un enrichissement « Gene Ontology » dans les intervalles de confiance de ces QTL va nous permettre de détecter les gènes surreprésentés dans leur voisinage (Monclus et al. 2012). De même, la localisation, dans les intervalles de confiance des QTL, des gènes candidats significatifs dans les études d'associations déjà disponibles sur la qualité du bois du peuplier (Wegrzyn et al. 2010; Guerra et al. 2013) pourrait révéler les candidats les plus prometteurs car à la fois positionnels, fonctionnels et/ou expressionnels et significatifs en association.

Chapitre VI

- Discussion générale et perspectives

Chapitre VI : Discussion générale et perspectives

La compréhension du déterminisme génétique des caractères d'intérêt pour les TCR et TtCR représente un enjeu important pour l'amélioration de la ressource lignocellulosique. Notre objectif à long terme est l'amélioration génétique des individus *P. nigra* pour la sélection des hybrides dédiés pour la production d'une biomasse forestière à travers une culture TCR ou TtCR. Actuellement, la ressource commence à être largement étudiée en vue de sa conversion en bioéthanol. On assiste à des efforts réels et non pas de mode chez différentes espèces pour une application industrielle vu le contexte actuel de la raréfaction des ressources pétrolières (Wegrzyn et al. 2010; Brereton et al. 2010; Studer et al. 2011; Wang et al. 2012; Ray et al. 2012; Jagtap et al. 2012; Guerra et al. 2013).

Cette thèse a permis d'explorer la piste moléculaire chez l'espèce *P. nigra* encore peu étudiée à ce niveau. La valorisation du potentiel génétique de cette espèce autochtone est un défi de par son intérêt en hybridation. Cette thèse a tout d'abord permis d'estimer le niveau de diversité nucléotidique de 456 gènes dans un set de 21 individus échantillonnés dans des populations naturelles françaises représentant la variabilité européenne de cette espèce (Chapitre II). Elle a conduit à la construction des cartes génétiques associées à une famille de plein-frères intraspécifique (Chapitre III). Elle a fourni, pour la première fois chez le peuplier, une évaluation de la variabilité génétique mais aussi de la covariation de caractères d'intérêt pour les TCR et TtCR comme la résistance à la rouille foliaire et la qualité du bois avant de détecter des QTL contrôlant cette variation (Chapitres IV et V).

A. Méthodes de découvertes de polymorphisme

Au début de la thèse, plusieurs stratégies étaient disponibles pour identifier du polymorphisme nucléotidique : l'*ecotilling* (Comai et al. 2004), le séquençage avec la méthode *Sanger*, et le pyroséquençage. Notre choix s'est porté sur un séquençage avec la méthode *Sanger* de fragments de gènes candidats sur un panel découverte de *P. nigra* d'origine géographique différente car cette méthode permet à la fois de détecter du polymorphisme SNP et d'obtenir une information de séquence généralement de bonne qualité. Parmi les deux autres stratégies, l'*ecotilling* est une méthode peu coûteuse et rapide mais elle ne permet pas d'accéder à toute l'information de séquence. Le pyroséquençage a quant à lui un débit élevé mais le pourcentage d'erreur est aussi plus important que celui du séquençage avec la méthode *Sanger*.

Les techniques de séquençage et génotypage sont en développement hallucinant. Il est aujourd'hui possible d'envisager le reséquençage de plusieurs individus pour la découverte de

polymorphisme voire leur génotypage (Huang et al. 2010 ; Cao et al. 2011; Slavov et al. 2012). Toutefois, l'assemblage des séquences *de novo* est actuellement en phase exploratoire et nécessite une infrastructure en bio-informatique très importante pour le traitement de données de séquences. L'assemblage peut également se faire par ancrage sur la séquence d'un génome de référence (si elle est disponible), ce qui n'est pas forcément une copie complète de l'espèce étudiée. À titre d'exemple, seulement 80% de séquences *de novo* de 52 individus *P. nigra* sont ancrés sur le génome de référence *P. trichocarpa* (Noveltree 2012). Outre le problème de l'assemblage, la qualité des SNP définis à partir des séquences issues de séquençage NGS (Nouvelle génération de séquençage) va dépendre de la qualité de la séquence et de la profondeur du séquençage. D'ailleurs, ce sont les deux critères de choix pour assainir le jeu de données SNP. En matière de génotypage, c'est la technique SNPLex qui a été utilisée dans le cadre de cette thèse. En 2005, cette méthode était considérée comme la méthode de haut débit par excellence (Tobler et al. 2005). Actuellement, la société *Applied biosystem* a abandonné cette technologie et ne la développe plus, soulignant la rapidité des progrès effectués dans ce domaine. En effet, le génotypage se fait désormais principalement au moyen de puces SNP dont la plupart sont préalablement identifiés par reséquençage NGS de plusieurs génomes (Slavov et al. 2012; Noveltree 2012) avec deux contraintes majeures l'une technique et l'autre budgétaire. Le génotypage par reséquençage (GBS) constitue le dernier cri des techniques de génotypage (Elshire et al. 2011). Les contraintes de cette technique sont les mêmes que pour les NGS. Pour les surmonter, on peut réduire la complexité du génome avec par exemple des enzymes de restriction (RADseq) ou de la capture de séquence codante (exome sequencing). Chez les plantes, deux études d'association sur le riz, dont le génotypage est réalisé par reséquençage de génomes entiers, ont récemment été publiées (Huang et al. 2010; Huang et al. 2012). Ces travaux soulignent l'intérêt de ces approches de génotypage à haut-débit pour entreprendre des études d'association sur tout le génome.

B. Analyse de ce polymorphisme nucléotidique

1. Intérêts et perspectives de l'analyse de la seule diversité moléculaire

L'analyse du polymorphisme SNP dans plusieurs gènes candidats dans un panel découverte de 21 *P. nigra* nous a permis d'estimer la diversité nucléotidique chez *P. nigra* qui est comparable à celle observée chez d'autres espèces de peuplier, chez des espèces forestières plus distantes et chez les plantes en général (Savolainen and Pyhajarvi 2007). La prochaine étape de notre analyse va s'attacher à déterminer dans les gènes candidats étudiés quelle est la part de la diversité silencieuse par rapport à la diversité fonctionnelle. La diversité silencieuse est

traditionnellement définie comme la somme de la diversité synonyme (n'impliquant pas de changement d'acide aminé) et de la diversité non codante tandis que la diversité fonctionnelle correspond à la diversité non synonyme c'est-à-dire impliquant un changement d'acide aminé. Ces deux définitions sont discutables. En effet, les polymorphismes dans les régions régulatrices sont classés dans la diversité silencieuse alors qu'ils peuvent avoir de forts effets sur les phénotypes et de fait ne sont pas réellement silencieux. *A contrario*, certains changements d'acide aminé peuvent ne pas avoir d'effet sur les phénotypes car, d'une part ils peuvent être éloignés du site actif de la protéine et d'autre part, ils peuvent avoir des propriétés physico-chimiques similaires (par exemple des changements à l'intérieur de la classe des acides aminés aliphatiques).

L'estimation de la diversité silencieuse donne une idée du taux de polymorphisme chez l'espèce étudiée tandis que la diversité fonctionnelle, généralement plus rare, nous informe directement de la fonction de l'allèle. Il existe d'autres formes de diversité de séquences. Les données relatives aux évènements d'indel, au nombre de copies de gènes (CNV), ou encore aux zones de méthylation de l'ADN seraient aussi des informations intéressantes à explorer. En effet, les indel sont souvent écartés pour le calcul des indices de diversité alors qu'ils représentent une source importante de la variabilité. En effet, le reséquençage de 80 accessions d'*Arabidopsis thaliana* a révélé près de 5 millions de SNP mais aussi plus de 800 000 indel courts et plus de 170 000 variants de structure (Cao et al. 2011). En ce qui concerne les CNV, le reséquençage de tétrade lors de la méiose chez des mutants *d'Arabidopsis thaliana* a révélé que la méiose produit ce type de variants distribués différemment dans le génome en fonction des croisements effectués (Lu et al. 2012). Dans d'autres cas, la variabilité de CNV permettrait probablement à la plante de renforcer ces défenses en ayant plusieurs copies de gènes de défense. En effet, chez le pin sylvestre, il a été démontré une variabilité du nombre de copies du gène antifongique (thaumatin-like) qui pourrait impliquer une variabilité de la résistance aux pathogènes (Skipars et al. 2010). Concernant la méthylation de l'ADN, on a mis en évidence un pattern de méthylation différent chez le pin *radiata* selon l'âge des arbres et leur état de convalescence après un stress (Fraga et al. 2002). Cette variation du niveau de méthylation peut jouer un rôle important dans le développement et la croissance des conifères ayant conservé un taux de polyploïdie élevé contrôlé par ces patterns épigénétiques (Grattapaglia et al. 2009). L'implication du contrôle de l'épigénétique dans les mécanismes de défenses des plantes et de leur interactions avec les agents pathogènes est par ailleurs décrit dans la littérature (Alvarez et al. 2010; Boyko and Kovalchuk 2011). En outre, une étude chez 6 génotypes de peuplier hybride a montré une corrélation positive entre le niveau de méthylation de l'ADN et la production de biomasse en

condition hydrique favorable (Gourcilleau et al. 2010). Ces travaux soulignent l'importance potentielle de ces polymorphismes qu'il conviendrait d'analyser pour une estimation plus complète de la diversité présente dans les populations.

L'étude du polymorphisme de séquences nous a fourni une nouvelle estimation du DL chez l'espèce *P. nigra.* Celui-ci chute rapidement au-delà de 400 pb comme chez la plupart des espèces du genre *Populus* (Ingvarsson 2005b; Chu et al. 2009; Wegrzyn et al. 2010; Guerra et al. 2013; Ismail et al. 2012; Kelleher et al. 2012), mais aussi chez des conifères (Gonzalez-Martinez et al. 2006). Outre son intérêt pour la détection de QTL, la génétique d'association et la sélection assistée par marqueurs, le déséquilibre de liaison peut être utilisé pour détecter des signatures de sélection. C'est le cas notamment lors de balayages sélectifs induit par la sélection de mutations bénéfiques (Kim and Nielsen 2004).

2. Un phénotypage de qualité qui reste encore coûteux

L'estimation de la valeur du phénotype étudié est très importante pour le succès des études d'association (Atwell et al. 2010), et plus généralement la détection de QTL et l'amélioration génétique. Deux contraintes de taille s'imposent : la première est la qualité de la mesure, la deuxième est le haut débit. Le développement du phénotypage à haut débit est un défi majeur que rencontre tous les chercheurs s'intéressant à des fonctions biologiques complexes. Ce développement n'a pas suivi l'envolée des méthodes de séquençage et de génotypage. Les méthodes de phénotypage de la résistance à la rouille abordées dans la thèse sont lourdes à mettre en œuvre et demandent un investissement humain important pour la mise en place des tests d'inoculation et la réalisation des observations sur une période continue de 10 jours. Elles mériteraient également une plus grande homogénéisation de l'inoculation sur de grands dispositifs. L'utilisation de ces méthodes nous a cependant permis d'évaluer les paramètres génétiques de différentes composantes épidémiologiques de la résistance partielle à 2 souches de *Mlp* dans un pédigrée de cartographie et de détecter des QTL pour ces caractères.

En ce qui concerne la qualité du bois, les méthodes de phénotypage directes et indirectes n'étaient pas utilisées en routine dans notre laboratoire avant ce travail de thèse. Un important investissement dans leur optimisation et dans l'augmentation des débits a donc été investi. Grâce à la SPIR, il a été possible d'étudier plusieurs composantes chimiques du bois ainsi que le potentiel de saccharification. Nos modèles de calibration sont de qualité comparable à ceux présents dans la littérature. Ce phénotypage indirect a permis aussi de mettre en évidence une gamme de variabilité génétique élevée pour les caractères de qualité du bois et de déceler des QTL contrôlant cette variabilité. Les modèles de phénotypage développés au cours de cette thèse

constituent une base pour établir des modèles de calibration plus généraux pour prédire une gamme de variabilité plus large que dans une famille de plein-frères, en s'intéressant aux génotypes du plan factoriel 5x4 et de la population d'association. Pour cela, la stratégie retenue sera de prédire les propriétés du bois de tous les échantillons avec les modèles actuels, de sélectionner quelques échantillons bien répartis sur la gamme de variation d'un caractère représentatif (comme la teneur en lignine) afin d'effectuer de nouveaux dosages chimiques et d'améliorer les modèles existant en introduisant ces nouvelles données. La prédiction par la SPIR est une méthode de phénotypage très intéressante et prometteuse mais la préparation des échantillons pour l'analyse reste un travail colossal. La récolte des échantillons n'est d'abord pas triviale : l'âge des arbres, le choix de l'échantillonnage intra-cépée, l'entretien de la parcelle (étiquetage et mauvaises herbes) et les conditions météorologiques rendent cette première étape compliquée. Après la récolte, les échantillons doivent être transportés, étiquetés, éventuellement écorcés, stockés dans un espace dédié et séchés convenablement (sinon ils risquent de s'abimer, cette étape dure en moyenne 15 jours). Ces échantillons sont ensuite découpés (environ 3 min par échantillon) puis broyés (environ 5 min par échantillon) avant d'être tamisés pour les analyses chimiques (environ 3 min par échantillon). Pour le potentiel de saccharification, une étape de préparation supplémentaire consiste à microniser les poudres (granulométrie < 50 µm) au moyen d'un broyeur à billes. Au total, il a fallu un mois de travail à temps plein pour la seule préparation des 479 échantillons du pédigrée. En vue d'améliorer le débit de ces étapes un broyage automatique et robotisé pourrait être envisagé (www.labman.co.uk/BioFuels.html).

3. Vers la génétique d'association ?

Le faible niveau de DL caractérisant les arbres forestiers rend les études de génétique d'association très précises mais présente l'inconvénient de nécessiter une densité de marquage énorme si on veut être exhaustif sur tout le génome. À titre d'exemple, si on considère qu'un génome de peuplier fait environ 485 Mb, que le niveau de DL (estimé avec le r^2) chute à 0,2 au bout de 400 pb et qu'il est uniforme tout au long du génome (ce qui est faux), il faudrait plus d'1,2 millions de SNP pour couvrir tout le génome et s'assurer une détection exhaustive en génétique d'association. Malgré la forte diminution des coûts de séquençage et de génotypage, cela n'est pas encore envisageable sur une population d'association (plusieurs centaines d'individus) chez le peuplier. Pour contourner ce problème, différentes approches peuvent être envisagées : 1/ des études d'associations dans des gènes candidats avec une contrainte majeure qui est le choix de ces gènes ; 2/ l'utilisation de populations caractérisées par un niveau de déséquilibre de liaison plus élevé comme des pédigrées complexes.

a) *Le choix des « gènes candidats » pour la génétique d'association :*

Différents types de candidats peuvent être considérés : 1/ des candidats fonctionnels, c'est à dire dont la fonction a été démontrée comme contrôlant le caractère d'intérêt soit dans la même espèce soit dans une espèce différente ; 2/ des candidats expressionnels, c'est-à-dire dont le niveau d'expression varie avec le caractère d'intérêt ; 3/ des candidats positionnels, c'est-à-dire qu'ils sont localisés dans une région génomique correspondant à un QTL associé à la variation du caractère d'intérêt. L'appartenance à une catégorie n'est pas exclusive, les candidats à la fois positionnels, fonctionnels et expressionnels représentant les candidats les plus intéressants.

Chez le pin maritime, Lepoittevin (2009) a choisi des facteurs de transcription (candidats à la fois expressionnels et fonctionnels) impliqués dans la formation du bois pour effectuer une étude d'association entre les polymorphismes de ces gènes et la croissance ainsi que les composantes de la qualité du bois (Lepoittevin 2009). Peu de SNP étaient significatifs dans les tests d'association. Ces gènes candidats étaient peu polymorphes, ce qui est probablement dû à une forte sélection purifiante, et ce qui a réduit le nombre des tests d'association avec le caractère étudié. Une façon de pallier ce problème, pourrait être de choisir des candidats en aval de voies de biosynthèse. En effet, il a été récemment montré sur *P. tremuloides* une corrélation positive entre le niveau de diversité et la position dans la voie pour des gènes de la voie de biosynthèse des lignines (Kelleher et al. 2012).

Chez *P. nigra*, 40 gènes candidats fonctionnels ont été analysés en génétique d'association avec des caractères de la qualité du bois. Six SNP correspondant à seulement 2 gènes étaient significativement associés avec ces caractères (Guerra et al. 2013). Chez *P. trichocarpa*, un total de 37 associations dans 16 gènes candidats, parmi le même set de gènes que Guerra et al. (2013), ont été détectées avec les caractères de la composition du bois (Wegrzyn et al. 2010). Toutefois ces associations n'expliquaient qu'une très faible part de la variation phénotypique (entre 1,19 et 3,84%). De même, Thumma et al. (2005) ont détecté une association significative entre le gène CCR et l'angle des microfibrilles du bois chez l'eucalyptus, mais qui n'explique que 5,9% de la variation phénotypique.

Le faible nombre d'associations détectées d'une part, et d'autre part le faible pourcentage de variation expliquée par les associations significatives, nous conduisent à nous interroger sur la validité de la démarche gène candidat. D'abord, si on considère que le choix des candidats était pertinent, on pourrait expliquer ces résultats par le fait que la variabilité dans les gènes choisis et dans les populations d'étude n'était pas suffisante pour permettre l'identification d'associations importantes. Ensuite, il se peut que les polymorphismes causaux les plus importants n'existaient pas dans le set d'identification des SNP. En effet, pour certains caractères le choix peut

apparaître très difficile au vue de la complexité des voies métaboliques correspondantes. C'est typiquement le cas des voies de biosynthèses impliquées dans la formation du bois (Zhong and Ye 2009 ; 2010 ; McCarthy et al. 2010 ; Du and Groover 2010 ; Drost et al. 2010 ; Dharmawardhana et al. 2010 ; Demura and Ye 2010). Enfin, les effets d'interaction entre gènes ne sont généralement pas capturés en génétique d'association sur gènes candidats, puisque d'une part ils ne sont généralement pas testés et que d'autre part les candidats choisis n'impliquent pas nécessairement des partenaires métaboliques susceptibles d'interagir.

Pour prendre en compte l'épistasie, d'autres catégories de gènes comme les miARN pourraient présenter un réel intérêt puisqu'ils sont généralement considérés comme des super-régulateurs des voies de biosynthèse. Les miARN interviennent dans les processus de développement de la plante et de sa croissance, ou en réponse à différents stress, ou dans les mécanismes de défense chez le peuplier (Lu et al. 2005; Lu et al. 2008). La présence de nombreux miARN chez le peuplier qui n'ont pas d'analogues chez *Arabidopsis thaliana* (Barakat et al. 2007) suggère aussi leur probable implication dans des processus spécifiques à la biologie des arbres. Par exemple, les miARN interviennent dans la formation du bois de tension (Lu et al. 2005). Il est aussi intéressant de noter que chez l'épicéa, les miARN co-localisent avec des clusters de NB LRR (Noveltree 2012). Les auteurs ont émis l'hypothèse que ces miARN se font détourner au profit de la machinerie cellulaire du pathogène qui secrètent des éliciteurs capables de les activer et éteindre grâce à eux l'action des gènes de défense NB LRR. Ces gènes sont donc des candidats intéressants pour les caractères étudiés dans ce travail de thèse. En vue d'aller plus loin dans leur analyse, une étude préliminaire a été entreprise sur les données de reséquençage présentées dans le chapitre II. Parmi les 665 gènes reséquencés, 26 étaient des cibles potentielles *in silico* des miARN. Parmi ces 26 gènes cibles, 2 présentaient notamment une homologie parfaite avec les 21 nucléotides du miRNA. Il s'agit de gènes codant pour une phospholipase et une sulfate adényltransférase. Leur fragments dans le panel de 21 *P. nigra* faisaient 640 et 590 pb et contenaient 8 et 7 SNP respectivement. Aucun SNP n'a été détecté dans le site de fixation du miARN.

Pour faciliter le tri de la multitude de candidats fonctionnels et expressionnels, il est aussi possible d'utiliser les informations de cartographie de QTL qui fournissent des candidats positionnels. En effet, la taille de l'intervalle de confiance d'un QTL responsable de la couleur de la baie chez la vigne a pu être réduite en combinant des informations fonctionnelles et positionnelles. Le polymorphisme des gènes candidats ainsi sélectionnés a permis de révéler une association qui explique 84 % de la variation phénotypique de la couleur de la baie (Fournier-Level et al. 2009). De même, l'analyse de la diversité d'un gène candidat positionnel et

fonctionnel chez l'eucalyptus a permis de cartographier finement un polymorphisme contrôlant la teneur en cellulose (Thumma et al. 2009). Cependant, d'autres études ont échoué dans la détection des associations entre des candidats positionnels et le caractère de la phénologie (Alberto 2010) ceci a été du probablement à des problèmes de structuration de la population d'association.

b) *L'étude de croisements multiples : une alternative plus réaliste à la génétique d'association sur tout le génome chez les arbres forestiers ?*

Parmi les alternatives aux études d'association dans les populations naturelles où le DL chute très rapidement et qui nécessitent un nombre très important de marqueurs, des associations dans des fonds génétiques moins hétérogènes peuvent être envisagées. En effet, le DL dans des populations interconnectées ou encore dans un plan factoriel est généralement plus élevé que dans les populations d'association traditionnellement utilisées (de type core-collections). Cette approche apparait ainsi plus réaliste en vue d'être plus exhaustif sur le génome. Toutefois, sa limite réside dans le faible nombre d'allèles étudié qui reste tout de même plus important que les 4 classes alléliques qu'on étudie au maximum dans les croisements biparentaux. Par ailleurs, ce type de dispositif pourrait aussi permettre de valider les effets des QTL détectés en croisements biparentaux dans des fonds génétiques différents (via un plan de croisement factoriel par exemple). L'utilisation de cette approche chez le palmier à huile a notamment permis de démontrer son intérêt par rapport à une approche biparentale (Billotte et al. 2010). Ainsi l'utilisation des croisements effectués dans les programmes d'amélioration constitue une voie prometteuse pour l'identification de régions génomiques contrôlant les caractères d'intérêt. Toutefois, à moins d'augmenter fortement le nombre de descendants, cette approche ne résout pas le problème de l'identification des polymorphismes causaux sous-jacents.

c) *Mais alors comment trouver les gènes sous-jacents et les polymorphismes causaux?*

Pour identifier les gènes sous-jacents des QTL, on peut cartographier finement les QTL à la recherche du gène ou du cluster de gènes qui contrôle le caractère. Cependant, il faut déployer des moyens considérables pour générer un nombre très important de descendants, les planter et les entretenir. Il a fallu par exemple générer 1500 descendants d'un pédigrée interspécifique (*P. deltoides x P. trichocarpa*) pour cartographier finement le locus « R_{US} » qui confère une résistance vis-à-vis de *Mlp* en réduisant la taille des urédosores du champignon (Bresson et al. 2011). Avec l'avènement des NGS, le reséquençage de pools d'individus extrêmes a récemment été proposé et utilisé avec succès chez la levure (Ehrenreich et al. 2010). Cette approche

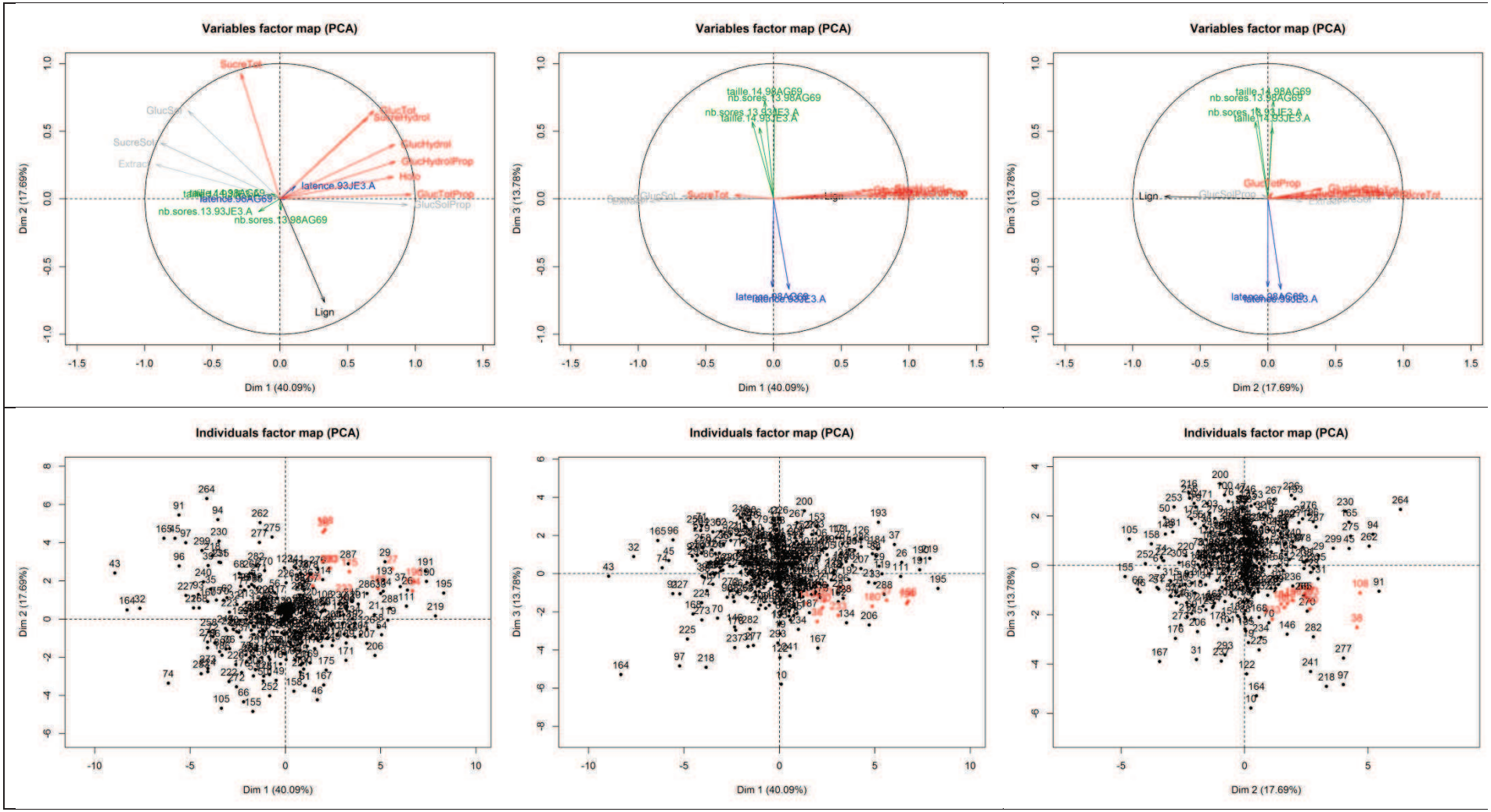


Figure VI. 1 : Représentation graphique de l'analyse en composante principale effectuée sur les moyennes clonales des caractères de la résistance à la rouille foliaire et la qualité du bois (La latence en bleu et nombre et taille des sores en vert. Les teneurs en lignines en noir, en solubles en gris et en holocellulose et sucres totaux et hydrolysés en rouge). Les graphiques du haut représentent les cercles de corrélation des variables sur les 3 premiers axes de l'ACP, tandis que ceux du bas représente la projection des génotypes sur ces mêmes axes. Les individus sélectionnés à partir de leur coordonnées sur les axes de l'ACP (coordonnée supérieure 1 sur l'axe 1, supérieure à 1 sur l'axe 2 et inférieure à -1 sur l'axe 3) sont indiqués en rouge.

consiste à identifier des extrêmes dans de larges descendances où ségrégent un QTL d'intérêt et de les séquencer par pool avec les NGS. Cette stratégie s'apparente à la « Bulk Segregant Analysis » (BSA ; (Michelmore et al. 1991)). Les marqueurs dont les fréquences alléliques différencient les deux groupes peuvent alors être identifiés. En perspective de mon travail, cette stratégie pourrait être envisagée pour réduire l'intervalle de confiance des QTL identifiés et notamment le QTL à large effet détecté pour la résistance à la rouille foliaire et localisé sur le chromosome 1. Pour cela, plutôt que de générer de nouveaux descendants, il est possible de chercher dans le programme d'amélioration génétique du peuplier des descendants du parent BDG porteur de ce QTL. De fait, cet individu avait déjà été repéré comme étant un bon parent dans le programme d'amélioration. Près d'un millier de descendants ont pu être recensés. Il suffirait maintenant d'identifier, parmi tous ces descendants, des extrêmes pour la résistance à la rouille foliaire et les séquencer en pools.

C. Deux caractères ? Deux défis ? Deux déterminismes différents ?

En ce qui concerne les perspectives d'amélioration génétique du peuplier pour une culture en TCR et TtCR, les analyses de variabilité génétique des caractères de résistance partielle à la rouille foliaire et de qualité du bois sont encourageantes puisque des niveaux d'héritabilité et une gamme de variabilité génétique relativement élevés ont été observés. Toutefois, dans le cas de la résistance à la rouille, il conviendrait de valider ces résultats d'une part avec un nombre plus important de souches de pathogène et d'autre part en conditions d'infection naturelle sur plusieurs cycles de l'agent pathogène. Dans le cas de la qualité du bois, il conviendrait aussi de valider les résultats chez des arbres plus âgés pour lesquels la proportion de bois analysé par rapport à l'écorce serait plus importante que dans le cadre de ce travail.

La sélection de génotypes intéressants pour les TCR et TtCR combinerait les deux caractères à savoir des individus résistants, ayant des teneurs élevées en sucres et des teneurs faibles en lignine. Pour cela il est nécessaire d'étudier la covariation de ces caractères afin de s'assurer de leur indépendance. En effet, il serait possible qu'une corrélation positive existe entre le taux de lignine et la résistance à la rouille, puisqu'il a été montré que la biosynthèse des lignines est activée en réponse à des stress et notamment à l'infection de pathogènes (Vanholme et al. 2010). Pour vérifier cela, nous avons effectué une analyse des corrélations entre les moyennes clonales des deux jeux de caractères (au moins 251 génotypes ont toutes les valeurs). Cette analyse a révélé des niveaux de corrélation très faibles entre les caractères de résistance à la rouille et les propriétés du bois. En effet, la corrélation la plus élevée en valeur absolue était de -0,15 entre le nombre de sores produits en réponse à l'infection avec la souche « 93JE3 » et la

Tableau VI.1 : Performances moyennes pour des caractères de résistance à la rouille foliaire et de propriétés du bois de l'ensemble du pédigrée et d'un sous-ensemble de génotypes sélectionnés. À partir de ces moyennes les gains génétiques ont été estimés pour tous les caractères. Les génotypes ont été sélectionnés sur la base de leurs coordonnées dans une analyse en composantes principales effectuée sur tous les caractères (cf. Figure VI.1).

Caractère	Moyenne des génotypes sélectionnés	Moyenne du pédigrée	Gain génétique (%)
Résistance à la rouille			
Souche 98AG69			
Latence	7.45	7.23	3.02
Nb. Sores	22.44	27.14	-17.32
Taille Sores	2.80	3.62	-22.61
Souche 93JE3			
Latence	9.38	8.56	9.53
Nb. Sores	8.75	16.30	-46.34
Taille Sores	2.55	2.97	-14.05
Propriétés du bois			
Dosages chimiques			
Extractibles (%)	7.74	9.31	-16.78
Lignines (%)	24.73	25.36	-2.47
Holocellulose (%)	68.29	65.74	3.89
Solubles (avant hydrolyse)			
Sucres [µmol/mg]	0.13	0.17	-19.69
Glucose [µmol/mg]	0.07	0.08	-3.10
Proportion du Glucose	0.59	0.49	22.54
Totaux (après hydrolyse)			
Sucres [µmol/mg]	0.51	0.46	9.76
Glucose [µmol/mg]	0.33	0.25	33.39
Proportion du Glucose	0.60	0.52	15.93
Hydrolysés (Totaux – Solubles)			
Sucres [µmol/mg]	0.39	0.31	25.62
Glucose [µmol/mg]	0.25	0.18	41.64
Proportion du Glucose	0.63	0.55	14.35

proportion de glucose hydrolysé. Ces résultats soulignent une indépendance de ces caractères dans le pédigrée étudié et de fait des QTL différents ont été identifiés. Ainsi une sélection multi-caractère pourrait être envisageable sur la base de ces résultats.

Pour cela nous avons effectué une analyse en composantes principales des moyennes clonales de tous les caractères (**Figure VI.1**). Les trois premiers axes de variation expliquent plus de 70% de la gamme de variation des moyennes clonales et les variables s'organisent sur ces axes de la façon suivante : 1/ sur l'axe 1, on retrouve principalement les teneurs en sucres et le glucose hydrolysé, la teneur en holocellulose, la proportion de glucose dans les sucres et le glucose total en valeurs positives et les sucres solubles et les extractibles en valeurs négatives ; 2/ sur l'axe 2, on retrouve les sucres totaux pour les valeurs positives et la teneur en lignines pour les valeurs négatives ; 3/ sur l'axe 3, on retrouve les variables de résistance à la rouille avec pour les valeurs positives les nombre et taille des sores et pour les valeurs négatives les durées de latence. Sur la base de ces résultats, une sélection de 11 génotypes a été entreprise à partir de leurs coordonnées sur les axes d'ACP (indiqués en rouge dans les graphiques du bas de la **figure VI.1**), c'est-à-dire que nous avons sélectionné les génotypes dont les coordonnées sur les axes 1, 2 et 3 étaient respectivement supérieures à 1, supérieures à 1 et inférieures à 1. Les moyennes de tous les caractères estimées sur ces 11 génotypes sont présentées dans le **tableau VI.1** avec les moyennes estimées dans l'ensemble du pédigrée ainsi que les gains génétiques correspondants. Avec cette stratégie, des gains génétiques importants (entre 14 et 46%) sont notamment attendus pour les nombre et taille des sores ainsi que pour les sucres et le glucose hydrolysés.

En ce qui concerne les perspectives d'utilisation des marqueurs moléculaires pour l'amélioration génétique, le déterminisme génétique de la résistance à la rouille offre une opportunité pour la sélection assistée par marqueurs qui permet de manipuler quelques QTL à effets majeurs. En revanche, le déterminisme complexe de la qualité du bois, présume une orientation vers d'autres méthodes pour l'amélioration génétique de ce caractère. La sélection génomique part du postulat initial que la variabilité génétique dépend d'un nombre infini de petit QTL à effet faible (modèle infinitésimal). Sans chercher à les identifier, elle cherche plutôt une estimation globale de la somme des effets de tous les QTL dans le génome (Meuwissen et al. 2001). De la même manière que pour la SPIR, cette étape d'apprentissage consiste en la construction d'équations de calibration qui permettent de corréler les valeurs du phénotype et ceux du génotype dans une population dite de référence. On peut ensuite appliquer ce modèle sur des données génotypiques pour prédire la valeur des phénotypes. Déjà implémentée sur plusieurs espèces animales, cette approche n'en est qu'à ses débuts sur les espèces forestières (Resende et al. 2012a; Resende et al. 2012b) et mériterait un développement sur le peuplier.

Annexe

Table IV.S1 : List of molecular marker tested and information on parental genotype, position on the reference genome, technical information (probe sequence), and resources used in the present paper.

C'est le tableau IV.S1 de l'article qui est soumis en supplément, il contient l'information des génotypes des deux parents pour les différents marqueurs utilisés pour la cartographie.

Marker name	Type	71072-501	BDG	Position on linkage map	Position on genetic map in Gaudet et al (2008)	Position on genetic map in Cervera et al (2001)	Position on genome *P. trichocarpa* (x 10^5)	Scaffold position on V 2.0 genome of *P. trichocarpa*	Scaffold position on V 1.1 genome of *P. trichocarpa*	Primers or probes used for genotyping SNP and PAMSA markers	V 1.1 gene model of SNP fragment	V 2.0 gene model of SNP fragment	Best Hit Arabidopsis of SNP fragment	Defline_Tair of SNP fragment	Defline V 2.0 *P. trichocarpa* of SNP fragment	Defline of other databases of SNP fragment
bi128189	SSR	145/149	145/151	XIII			131.22566	XIII	XIII							
bi135774	SSR	378/378	358/378	VI			229.29066	VI	VI							
bi138728	SSR	173/175	173/173	XIX			116.76333	XIX	XII							
bi139308	SSR	157/157	157/157				96.51422	VIII	VIII							
bi139327	SSR	NA	NA				13.41189	III								
bu810400	SSR	100/111	100/109	XIII	XIII		10.23091	XIII	XIII							
bu810907	SSR	395/395	395/395				39.68703	XIV	XIV							
bu813610	SSR	92/92	92/92				191.15668	I								
bu813833	SSR	164/AN	152/155	I	I		430.48192	I	I							
bu814260	SSR	113/113	113/113				88.43084	XIII	XIII							
bu814989	SSR	129/155	142/155	X	X		151.06363	X	X							
bu818855	SSR	195/195	192/198	XIII	XIII		44.83347	XIII	XIII							
bu831219	SSR	243/245	242/247	I			171.68514	I	I							
bu867968	SSR	NA	NA				68.82384	I								
bu885452	SSR	185/185	185/185		XVIII		125.49003	XVIII	XVIII							
bu890808	SSR	159/169	163/169	III			0.40283	scaffold_22	III							
GCPM_1007-1	SSR	153/AN	151/158				77.94459	IX								
GCPM_104	SSR	NA	NA				6.65009	XIX								
GCPM_1045-1	SSR	203/203	203/203				73.9918	III	III							
GCPM_1046	SSR	133/167	133/167					no hit	XVIII							
GCPM_107	SSR	199/199	193/199	XIX			4.19569	XIX	XIX							
GCPM_1101-1	SSR	NA	NA				108.39372	XII								
GCPM_1075-1	SSR	124/140	124/124	IV			192.94192	IV	IX							
GCPM_1112-1	SSR	NA	NA				120.73982	XVIII								
GCPM_1120-1	SSR	230/230	230/230				54.7665	X								
GCPM_1126-1	SSR	NA	NA				30.5586	XI								
GCPM_1131	SSR	NA	NA				64.25034	XIX	XIX							
GCPM_1133	SSR	187/187	187/187				47.16363	III	III							
GCPM_1186	SSR	251/251	240/240				138.72323	XII								
GCPM_1108-1	SSR	202/204	202/204	VII			36.86501	VII	VII							
GCPM_1220-1	SSR	209/225	217/225	XIII			86.21338	XIII	XIII							
GCPM_1224-2	SSR	156/156	156/156				110.39642	IX	IX							
GCPM_1223-1	SSR	134/140	140/140	XII			13.83782	XII	XII							
GCPM_1229-1	SSR	NA	NA				34.78598	VI								
GCPM_1231-1	SSR	233/252	252/AN				110.2998	IX								
GCPM_1235	SSR	NA	NA				114.91758	IV	IV							
GCPM_1225-1	SSR	160/AN	163/AN	V			71.52527	V	V							
GCPM_1253-1	SSR	NA	NA				102.4408	IX	IX							
GCPM_124	SSR	222/222	222/233	I			91.31218	I	I							
GCPM_1273	SSR	126/126	126/126				236.48447	VI								
GCPM_1292	SSR	189/191	191/193				80.54865	XIV	XIV							
GCPM_130-1	SSR	NA	NA				123.20686	XIII								
GCPM_1374-1	SSR	NA	NA				110.69232	XVI	XVI							
GCPM_1377-1	SSR	190/190	190/190				84.55977	XIV	XIV							
GCPM_1240-1	SSR	270/AN	240/250	XIII			129.52285	XVI	XVI							
GCPM_1384-1	SSR	NA	NA				143.62236	XVIII								
GCPM_1262-1	SSR	219/225	225/227	IX			111.55524	IX	IX							
GCPM_143-1	SSR	NA	NA				62.35041	IX								
GCPM_1263-1	SSR	247/251	247/253	I			404.78994	I	I							
GCPM_1438-1	SSR	NA	NA				22.29765	XVI	XVI							
GCPM_1460-2	SSR	NA	NA				151.22719	VIII								
GCPM_1381	SSR	257/269	248/269	XVII			61.00061	XVII	XVII							
GCPM_1525-1	SSR	224/224	224/224				15.26774	V								
GCPM_1418-1	SSR	205/208	205/205	XVI			3.51021	XVI	XVI							
GCPM_1547-1	SSR	NA	NA				14.02543	XVI	XVI							
GCPM_1554-1	SSR	NA	NA				137.20984	XVII								
GCPM_1589	SSR	NA	NA				161.29425	III								
GCPM_1433-1	SSR	169/174	164/176	XIV			131.15122	XIV	XIV							
GCPM_1604-1	SSR	230/230	230/230				126.13403	IX								
GCPM_161-1	SSR	170/170	170/170				120.43283	XIX								
GCPM_1629	SSR	NA	NA				113.10324	III								
GCPM_1649-1	SSR	NA	NA				143.6247	VII								
GCPM_146-1	SSR	225/230	175/217	XII			135.46869	XII	XII							
GCPM_1533-1	SSR	211/AN	215/238	XII			119.88113	XII	XII							
GCPM_1717-1	SSR	NA	NA				35.11612	XVI	XVI							
GCPM_174-1	SSR	NA	NA				11.3955	VI								
GCPM_1743	SSR	NA	NA				103.57086	XVIII	XVIII							
GCPM_1759-1	SSR	216/216	228/228				213.51161	V								
GCPM_1764-1	SSR	NA	NA				22.53199	XIX								
GCPM_187-1	SSR	NA	NA				34.35334	XV								
GCPM_1874-1	SSR	66/AN	66/75				120.89171	VI								
GCPM_1899-1	SSR	NA	NA				46.47852	XVI	XVI							
GCPM_1900-1	SSR	159/AN	160/AN				89.47484	XVI								
GCPM_1600-1	SSR	242/246	242/242	VIII			12.81064	VIII	VIII							
GCPM_1940	SSR	NA	NA				16.25299	XIV								
GCPM_20	SSR	NA	NA				231.71353	V								
GCPM_1653-1	SSR	308/308	288/308	XIX			94.68961	XIX	XIX							
GCPM_1699-3	SSR	206/215	206/249	XV			89.09966	XV	XV							
GCPM_2050-1	SSR	163/163	163/163				90.66077	XVI	XVI							
GCPM_2058-1	SSR	235/235	235/235				29.69416	IX								
GCPM_2080	SSR	225/225	225/225				103.462	XIV								
GCPM_2110-1	SSR	AN/AN	65/AN				115.8826	V								
GCPM_2123-1	SSR	NA	NA				33.36939	XVII								
GCPM_2155-2	SSR	NA	NA				37.39888	VII	VII							
GCPM_2180-1	SSR	NA	NA		XVI		84.4349	XVI	XVI							
GCPM_2197-1	SSR	NA	NA				10.89635	XVIII								
GCPM_2288-1	SSR	NA	NA				3.63309	VII	VII							
GCPM_2347	SSR	NA	NA				95.0842	X								
GCPM_2382-1	SSR	NA	NA				116.05456	XV								
GCPM_1719	SSR	220/230	220/246	I			354.88501	I	I							
GCPM_2425	SSR	215/215	215/215				4.9604	XI								
GCPM_2431	SSR	229/229	229/229				108.84447	X								
GCPM_2465	SSR	243/243	243/243				44.99587	XI								
GCPM_2486-1	SSR	NA	NA				13.54457	XV								
GCPM_2564-1	SSR	NA	NA				9.30394	XIX								
GCPM_1929-1	SSR	160/160	160/177	XIV	XIV		48.39226	XIV	XIV							
GCPM_2012-1	SSR	138/138	138/139	XIV			87.83021	XIV	XIV							

GCPM_2641-1	SSR	121/156	138/156				102.60759	XIII								
GCPM_2643	SSR	NA	NA					several hits	several hits							
GCPM_2662	SSR	NA	NA				105.67049	XVII	XVII							
GCPM_2729	SSR	NA	NA				81.90457	I								
GCPM_2015-1	SSR	287/299	266/299	I			370.60832	I	I							
GCPM_2755-1	SSR	NA	NA				49.51964	XV	XV							
GCPM_2779-1	SSR	NA	NA				112.29829	VI								
GCPM_278-1	SSR	126/126	126/126				54.60407	VII								
GCPM_2834	SSR	NA	NA				116.4578	XVII	XVII							
GCPM_2894	SSR	NA	217/236				21.81235	XIII	XIII							
GCPM_2024-1	SSR	227/229	229/229	XVI	XVI		132.74143	XVI	XVI							
GCPM_2951	SSR	175/175	175/175				2.20806	XI								
GCPM_2420-1	SSR	174/197	174/207	XIV			80.4021	XIV	XIV							
GCPM_3009-1	SSR	NA	NA				417.60519	I								
GCPM_2569-1	SSR	187/191	187/191	I			373.29558	I	I							
GCPM_2570-1	SSR	235/242	233/233	I			322.62783	I	I							
GCPM_3073-1	SSR	NA	NA				33.18718	XVI								
GCPM_3100-1	SSR	191/209	191/209				154.44728	XIX								
GCPM_3147-1	SSR	NA	NA				127.67467	XIV								
GCPM_3162	SSR	218/218	218/218				41.3628	XIX								
GCPM_3178-1	SSR	NA	NA				52.93619	XIX	XIX							
GCPM_2581-1	SSR	236/238	230/236	I			376.03839	I	I							
GCPM_3230-1	SSR	NA	NA				140.01025	XIV								
GCPM_2623-1	SSR	263/263	248/263	XIX			101.84248	XIX	XIX							
GCPM_3269	SSR	NA	NA				143.81091	III								
GCPM_3278	SSR	196/196	196/196				21.1487	XII								
GCPM_329-1	SSR	NA	NA				56.21771	XVIII								
GCPM_333-1	SSR	NA	NA				22.70208	I								
GCPM_3333	SSR	NA	NA				138.14063	XV								
GCPM_3345-1	SSR	NA	NA				137.53813	XVIII								
GCPM_3357	SSR	231/231	231/231				5.15059	XV								
GCPM_336-1	SSR	207/207	207/207				56.47406	XIV	XIV							
GCPM_3415-1	SSR	NA	NA				427.23979	I								
GCPM_342	SSR	NA	NA					several hits	several hits							
GCPM_1326-1	SSR	248/248	248/248				370.11466	I								
GCPM_2630-1	SSR	234/235	235/238	IX			83.05726	IX	IX							
GCPM_1683-1	SSR	112/132	138/138				231.22806	I								
GCPM_3486-1	SSR	NA	NA				28.58964	X								
GCPM_3496	SSR	217/AN	217/AN				7.77525	XVI								
GCPM_350-1	SSR	222/222	222/222		VII		145.08666	VII	VII							
GCPM_3503-1	SSR	NA	NA				108.9204	X	XVI							
GCPM_3504-1	SSR	252/AN	254/AN				69.08537	III	III							
GCPM_3523-1	SSR	NA	NA				71.97968	XII								
GCPM_3576-1	SSR	NA	NA				98.24935	XV								
GCPM_3671	SSR	NA	NA				18.11425	V								
GCPM_3675-1	SSR	170/170	170/170				94.88053	XVI	XVI							
GCPM_3680-1	SSR	164/164	164/164				65.32303	XIX								
GCPM_3687-1a	SSR	227/227	227/227				121.71712	IX	IX							
GCPM_3687-1b	SSR	227/227	227/227				121.81145	IX	IX							
GCPM_1704-1	SSR	210/210	210/210				337.28753	I								
GCPM_2743-1	SSR	153/153	150/153	XV			50.37295	XV	XV							
GCPM_3709-1	SSR	NA	NA				66.04954	XVI								
GCPM_3763	SSR	NA	NA				112.47587	XIX								
GCPM_3766	SSR	NA	NA				12.71379	XV	XV							
GCPM_3773	SSR	NA	NA				130.97812	III	III							
GCPM_3827-1	SSR	NA	NA				128.18061	IX	IX							
GCPM_386	SSR	NA	NA				13.23472	X								
GCPM_3871-1	SSR	241/241	241/241				33.24432	XIV								
GCPM_3873-1	SSR	230/230	230/230				88.80786	XVIII								
GCPM_2837	SSR	222/226	222/228	I			198.58245	I	I							
GCPM_2900	SSR	210/210	210/215	III			143.81091	III	III							
GCPM_3956-1	SSR	132/132	132/132				75.60263	X								
GCPM_3975-1	SSR	NA	NA				43.23479	V								
GCPM_3987-1	SSR	NA	NA				40.81971	XV	XV							
GCPM_2995-1	SSR	176/185	172/191	VII			117.78499	VII	VII							
GCPM_3029-1	SSR	241/245	241/245	VIII			145.79851	VIII	VIII							
GCPM_4063	SSR	NA	NA				58.19611	XII	XII							
GCPM_3050-1	SSR	229/229	222/229	XIX			51.91498	XIX	XIX							
GCPM_45-1	SSR	NA	NA				21.91045	IV	IV							
GCPM_465-1	SSR	NA	NA				117.81788	XIX	XIX							
GCPM_3208-1	SSR	179/183	179/179	VII			105.15085	VII	VII							
GCPM_3214	SSR	180/182	180/182	I			272.99512	I	I							
GCPM_512-2	SSR	NA	NA				445.36002	I								
GCPM_547-1	SSR	NA	NA				17.67861	VI								
GCPM_556-1	SSR	NA	NA				193.3813	III	III							
GCPM_557-1	SSR	NA	NA				30.09595	V								
GCPM_3243-1	SSR	258/262	262/277	I			444.77305	I	I							
GCPM_598-1	SSR	200/200	200/200				70.84495	XVI	XVI							
GCPM_636-2	SSR	220/220	220/220				112.70134	XVIII								
GCPM_641	SSR	NA	NA				131.54413	XVII	XVII							
GCPM_2439-1	SSR	176/176	176/176				246.69973	I								
GCPM_3476-1	SSR	208/214	202/214	VII			33.36365	VII	VII							
GCPM_3688	SSR	177/189	167/167	I			467.36752	I	I							
GCPM_37	SSR	172/172	172/174	VII			133.53661	VII	VII							
GCPM_748-1	SSR	NA	NA				166.65904	XIV	XIV							
GCPM_75-1	SSR	NA	NA				21.15379	XI								
GCPM_76-1	SSR	NA	NA				166.31696	VI								
GCPM_778-1	SSR	293/293	293/293				183.29443	VIII								
GCPM_3893-1	SSR	202/222	203/224	V			6.77174	VII	VII							
GCPM_801	SSR	NA	NA				76.33509	XVIII								
GCPM_818-1	SSR	NA	NA				341.77923	I								
GCPM_83-1	SSR	NA	NA				93.17267	III	III							
GCPM_2898-2	SSR	248/248	248/249				384.43344	I								
GCPM_856-1	SSR	127/127	127/127				40.64559	VII								
GCPM_3904-1	SSR	271/AN	235/257	VII			67.03462	VII	VII							
GCPM_4002-1	SSR	184/190	184/196	XIV			114.39362	XIV	XIV							
GCPM_3939	SSR	172/172	172/172				31.15806	I								
GCPM_833	SSR	230/230	230/230				23.02285	I								
GCPM_887	SSR	NA	NA				163.22613	III	III							
GCPM_895	SSR	NA	NA				168.225	II								

GCPM_923-1	SSR	NA	NA				145.24224	XVII								
GCPM_4051	SSR	203/221	204/214	XII			92.33401	XII	XII							
GCPM_446-1	SSR	241/249	240/240	XV			77.62201	XV	XV							
GCPM_966-1	SSR	NA	NA				148.79298	XIX								
GCPM_971-1	SSR	NA	NA				30.56193	XVIII								
GCPM_973-1	SSR	243/243	243/243				31.47156	IV	IV							
GCPM_985-1	SSR	145/145	145/145				161.59849	XI	XI							
GCPM_505-1	SSR	237/250	232/238	XVI			139.78645	XVI	XVI							
GCPM_512-1	SSR	233/236	210/233	I			445.35694	I	I							
GCPM_588	SSR	247/247	243/247	IX	IX		91.0074	IX	IX							
GCPM_712-1a	SSR	163/186	163/186	XIX			34.55739	XIX	XIX							
GCPM_712-1b	SSR	199/199	199/200	XIX				XIX	XIX							
GCPM_714-1	SSR	167/173	169/173	I			348.68254	I	I							
GCPM_787-1	SSR	141/155	155/155	II			171.89072	II	II							
GCPM_877-1a	SSR	113/113	113/140	I			363.23225	I	I							
GCPM_877-1b	SSR	219/219	219/259	I				I	I							
GCPM_943	SSR	208/208	206/208	XI	XI		169.4241	XI	XI							
GCPM_961	SSR	182/186	182/186	IV			10.71993	IV	IV							
ORPM_104	SSR	NA	NA				1.86951	VIII	VIII							
ORPM_106	SSR	NA	NA				92.28103	IV	IV							
ORPM_11	SSR	207/AN	AN/AN				46.19702	VIII	VIII							
ORPM_112	SSR	193/193	AN/AN				143.17632	VIII	VIII							
ORPM_114	SSR	225/225	225/225				34.56496	XVI	XVI							
ORPM_117	SSR	225/225	225/225				24.26422	II								
ORPM_118	SSR	NA	NA				52.39456	XVI								
ORPM_121	SSR	NA	NA				146.87416	VIII	XVI							
ORPM_122	SSR	217/217	217/217				24.59529	XVI								
ORPM_13	SSR	NA	NA				64.20508	XVIII								
ORPM_133	SSR	236/236	236/236				71.94769	XIV	XIV							
ORPM_134	SSR	223/223	223/223				0.6958	V								
ORPM_137	SSR	205/205	AN/AN				90.87505	I								
ORPM_14	SSR	NA	NA					several hits	several hits							
ORPM_144	SSR	212/212	212/212		IX		136.72267	VII	IX							
ORPM_147	SSR	256/256	256/256					several hits	several hits							
ORPM_148	SSR	197/197	197/197				56.61335	VI								
ORPM_15	SSR	NA	NA				58.4599	XVII	XVII							
ORPM_158	SSR	NA	NA				89.19163	XVIII	XVIII							
ORPM_166	SSR	225/225	225/225				138.26152	VII	VII							
ORPM_167	SSR	201/201	201/201				58.21054	VI	VI							
ORPM_173	SSR	NA	NA				473.56859	I								
ORPM_184	SSR	210/210	210/210				134.09838	V	V							
ORPM_186	SSR	NA	NA				59.8793	IV	IV							
ORPM_197	SSR	215/215	215/215				206.82403	VI	VI							
ORPM_20	SSR	220/220	220/220				61.22341	I								
ORPM_202	SSR	203/203	203/203				131.52048	VIII	VIII							
ORPM_105	SSR	216/218	216/218	V			58.98473	V	V							
ORPM_127	SSR	184/184	182/186	IV	IV		64.47111	IV	IV							
ORPM_209	SSR	228/228	228/228				46.46407	II								
ORPM_146	SSR	229/229	226/229	X			84.29192	X	X							
ORPM_210	SSR	220/220	220/220				83.46357	XII	XII							
ORPM_213	SSR	NA	NA				124.62552	XV	XV							
ORPM_214	SSR	196/198	196/198		XVIII		45.8397	XVIII								
ORPM_216	SSR	220/220	220/220				82.59151	VIII								
ORPM_217	SSR	228/228	228/228				137.87363	XI	XI							
ORPM_219	SSR	220/AN	220/AN				94.82102	VIII	VIII							
ORPM_220	SSR	240/240	240/240		IV			several hits	several hits							
ORPM_149	SSR	236/239	236/236	X			165.81673	X	X							
ORPM_226	SSR	239/239	239/239				1.98738	XIV								
ORPM_229	SSR	247/247	247/247				161.49297	XIV	XIV							
ORPM_23	SSR	NA	NA		IX		41.56608	IX	IX							
ORPM_230	SSR	NA	NA				82.18853	III	III							
ORPM_234	SSR	228/228	228/228				133.80211	XIX								
ORPM_235	SSR	225/225	225/225				32.55935	II								
ORPM_16	SSR	250/254	248/250	XIII	XIII (dupl.)		2.82451	scaffold 30	XIII							
ORPM_25	SSR	169/169	169/169				247.01992	V								
ORPM_252	SSR	207/207	207/207				98.97359	VII								
ORPM_257	SSR	227/227	227/227				149.00018	VII	VII							
ORPM_260	SSR	NA	NA				188.44287	II								
ORPM_263	SSR	NA	NA				31.16374	XIX	XIX							
ORPM_264	SSR	220/220	220/220		VIII		94.82102	VIII	VIII							
ORPM_268	SSR	217/217	217/217				134.26986	VIII	VIII / VI /XI							
ORPM_272	SSR	228/228	228/228				137.87363	XI								
ORPM_273	SSR	NA	NA				50.59049	IX								
ORPM_276 a	SSR	NA	NA		XIII		8.86814	XIX	XIX							
ORPM_276 b	SSR	NA	NA		XIII		28.50188	XIX	XIX							
ORPM_175	SSR	246/247	246/249	I			7.96329	I	I							
ORPM_176	SSR	213/225	225/225	XIV			55.46759	XIV	XIV							
ORPM_28	SSR	228/229	228/228				119.93406	XVIII	XVIII							
ORPM_285	SSR	221/221	221/221				72.23271	IX	IX							
ORPM_177	SSR	255/259	259/272	I	I		23.23928	I	scaffold_29							
ORPM_190	SSR	219/219	219/227	VI	VI		137.17996	VI	VI							
ORPM_292	SSR	238/238	238/238				1.17151	XVII								
ORPM_293	SSR	213/213	213/213				276.12117	I								
ORPM_295	SSR	226/226	226/226				36.95159	XIX	XIX							
ORPM_193	SSR	218/218	208/218	XIV	XIV		87.80658	XIV	XIV							
ORPM_201	SSR	230/230	230/233	IV			75.64874	IV	IV							
ORPM_301	SSR	224/224	224/224				107.5049	VIII								
ORPM_302	SSR	209/209	209/209				107.50144	VIII								
ORPM_303	SSR	235/235	235/235					several hits	several hits							
ORPM_304	SSR	220/220	220/220				123.34243	XI	XI							
ORPM_305	SSR	204/204	204/204				42.46676	VI								
ORPM_307	SSR	243/243	243/243				220.31272	II								
ORPM_310	SSR	263/263	263/263				103.61523	XV	XV							
ORPM_312	SSR	217/217	217/217				116.25105	VII	VII							
ORPM_313	SSR	237/237	237/237				76.62416	V	V							
ORPM_314	SSR	NA	NA				261.55132	I								
ORPM_203	SSR	232/234	234/234	III			120.09778	III	III							
ORPM_327	SSR	228/228	228/228				72.76518	VIII	VIII							
ORPM_33	SSR	220/220	220/220				146.81249	XI	XI							
ORPM_335	SSR	256/256	256/256				73.107	VIII								

ORPM_340	SSR	149/149	149/149				146.49296	X	X							
ORPM_206	SSR	211/216	216/216	XIX	XIX		36.30139	XIX	XIX							
ORPM_345	SSR	231/231	231/233				146.60481	X	X							
ORPM_347	SSR	NA	NA					several hits	several hits							
ORPM_367	SSR	NA	NA				229.17889	VI	VI							
ORPM_370	SSR	234/234	234/234				66.38147	VIII	VIII							
ORPM_372	SSR	212/212	212/212				65.50329	VIII	VIII							
ORPM_379	SSR	200/200	200/200					several hits	several hits							
ORPM_38	SSR	AN/AN	270/AN				65.32433	XVI	XVI							
ORPM_207	SSR	229/231	229/239	II			72.56919	II	II							
ORPM_393	SSR	251/251	251/251				4.93934	IV	IV							
ORPM_399	SSR	NA	NA		I		5.26464	IV	IV							
ORPM_40	SSR	198/198	198/198		II		173.2841	II								
ORPM_402	SSR	238/238	238/238					several hits	several hits							
ORPM_41	SSR	243/243	243/243				41.22795	XVIII								
ORPM_415	SSR	315/315	315/315				107.9146	III								
ORPM_420	SSR	235/235	235/235				15.26642	I								
ORPM_421	SSR	NA	NA				53.34456	IV	IV							
ORPM_21	SSR	234/235	237/238	IX			51.79738	IX	IX							
ORPM_427	SSR	197/197	197/197				88.85857	VIII	VIII							
ORPM_429	SSR	NA	NA				20.30258	XVIII								
ORPM_43	SSR	213/213	213/213				26.56921	VI	VI							
ORPM_432	SSR	220/220	220/220		XIX		126.97612	XIX	XIX							
ORPM_221	SSR	233/237	239/249	IV	IV		185.90672	IV	IV							
ORPM_438	SSR	234/234	234/234				41.97484	XV	XV							
ORPM_444	SSR	NA	NA		IX		41.5664	IX	IX							
ORPM_446	SSR	277/277	277/277				71.94769	XIV	XIV							
ORPM_447	SSR	221/221	221/221				90.87427	I								
ORPM_45	SSR	NA	NA		V		235.89512	V	V							
ORPM_451	SSR	226/245	226/245		IX		128.67828	IX	IX							
ORPM_455	SSR	NA	NA		VIII		111.14535	VIII	VIII							
ORPM_470	SSR	161/175	177/AN				120.09968	III	III							
ORPM_477	SSR	226/228	228/230		XV		39.16997	XVIII	XVIII							
ORPM_478	SSR	NA	NA				38.19755	XVIII	XVIII							
ORPM_240	SSR	252/291	262/274	I			189.97214	I	I							
ORPM_482	SSR	NA	157/160				38.55162	XVIII	XVIII							
ORPM_489	SSR	NA	NA				38.64422	XVIII	XVIII							
ORPM_496	SSR	NA	NA		XVIII		38.5914	XVIII	XVIII							
ORPM_50	SSR	220/AN	224/AN				260.61586	VI	VI							
ORPM_26	SSR	229/238	229/238	VI	VI		57.86927	no hit	VI							
ORPM_55	SSR	NA	NA				26.66338	XIII	XIII							
ORPM_269	SSR	254/256	256/AN	XI	VIII, XI		46.80897	XI	XI							
ORPM_59	SSR	237/237	237/237				27.17063	XIV	XIV							
ORPM_6	SSR	NA	NA				118.4992	XVIII	XVIII							
ORPM_277	SSR	200/211	200/AN	XIX			4.9	XIX	several hits							
ORPM_61	SSR	NA	NA				26.55186	VI								
ORPM_62	SSR	NA	NA				246.77023	I								
ORPM_64	SSR	249/249	249/249				17.2774	XVI	XVI							
ORPM_65	SSR	242/242	242/242				235.91825	V	V							
ORPM_66	SSR	237/237	237/237				13.25344	I								
ORPM_67	SSR	258/258	256/256				208.93442	IV	IV							
ORPM_279	SSR	234/242	246/AN	VI	VI		251.60369	VI	VI							
ORPM_73	SSR	NA	NA				143.06905	XV								
ORPM_75	SSR	NA	NA				286.83689	I								
ORPM_76	SSR	217/217	217/217				57.41393	IX	IX							
ORPM_80	SSR	204/204	204/204				18.43938	XII	XII							
ORPM_81	SSR	NA	NA				137.86501	XI	XI							
ORPM_86	SSR	230/230	230/230				32.52242	XVI	XVI							
ORPM_87	SSR	220/220	220/220				165.5273	IV	IV							
ORPM_88	SSR	213/213	213/213				34.56257	XVI	XVI							
ORPM_92	SSR	230/233	233/233				34.31936	VIII	VIII							
ORPM_95	SSR	256/256	256/256					several hits	several hits							
ORPM_287	SSR	262/264	264/AN	II	II		62.98346	II	II							
ORPM_29	SSR	257/259	271/281	XI			122.56418	XI	XI							
ORPM_297	SSR	234/240	246/253	XIII	XIII		154.29183	XIII	XIII							
ORPM_30	SSR	243/243	243/251	III	III		107.3806	III	III							
ORPM_32	SSR	274/278	278/282	XII			132.58962	XII	XII							
ORPM_344	SSR	263/275	263/275	X			147.38838	X	X							
ORPM_389	SSR	235/236	235/235	X			146.98454	X	X							
ORPM_425	SSR	195/198	195/198	IV			92.43703	IV	IV							
ORPM_435	SSR	230/231	231/231	VII			137.90931	VII	VII							
ORPM_479	SSR	258/258	255/258	XVIII			38.73904	XVIII	XVIII							
PMGC_108	SSR	NA	NA				97.97145	XII	XII							
PMGC_14	SSR	230/230	212/212			XIII	5.69201	XIII	XIII							
ORPM_488	SSR	217/217	217/218	VI			38.58041	XVIII	XVIII							
PMGC_2020	SSR	NA	NA			IX	206.59739	IV	IV							
PMGC_2030	SSR	NA	NA				9.60152	XVII	XVII							
PMGC_204	SSR	293/300	292/298		I			several hits	several hits							
PMGC_2055	SSR	NA	NA				77.2723	XIV	XIV							
PMGC_2060	SSR	140/140	140/140				39.69792	VIII	VIII							
PMGC_2084	SSR	NA	NA			NA	370.06106	I	no hit							
ORPM_54	SSR	242/245	242/245	I			286.8207	I	I							
PMGC_2098	SSR	NA	NA				289.07048	I								
PMGC_2105	SSR	115/115	115/115				94.24924	XV	XV							
ORPM_56	SSR	192/195	188/192	VIII	VIII		137.00789	VIII	VIII							
PMGC_2143	SSR	NA	NA				44.61985	XVI	XVI							
PMGC_2156	SSR	131/131	131/131		V	V		several hits	several hits							
ORPM_60	SSR	221/221	209/221	VI	VI		236.55254	VI	no hit							
ORPM_7	SSR	216/216	216/235	V			82.23043	V	V							
PMGC_223	SSR	150/162	150/162				161.75459	II								
ORPM_96	SSR	369/370	369/369	IX			106.56398	IX	IX							
PMGC_2274	SSR	123/123	116/126				174.52692	III	III							
PMGC_2321	SSR	318/318	318/318				38.86626	IX	IX							
PMGC_2392	SSR	NA	NA				84.78409	XI	no hit							
PMGC_2408	SSR	NA	NA				69.88092	XV	XV							
PMGC_2418	SSR	NA	NA				209.87628	II	no hit							
PMGC_2420	SSR	NA	NA				118.95842	XIII	XIII							
PMGC_244	SSR	NA	NA				22.10911	IX	IX							
PMGC_2481	SSR	NA	NA		III		161.16558	III	III							
PMGC_2499	SSR	NA	NA		I		43.10946	I								

PMGC_2522	SSR	173/173	173/173				45.68085	IX	IX							
PMGC_2531	SSR	NA	NA				172.346	XI	XI							
PMGC_2541	SSR	76/76	76/76				138.135	VI	VI							
PMGC_2011	SSR	102/115	100/115	XI	XI	XI	51.20502	XI	scaffold_3889							
PMGC_2558	SSR	153/153	155/155		V		200.11188	V	V							
PMGC_2571	SSR	174/174	174/174				122.2365	X	X							
PMGC_2088	SSR	190/197	180/188	II	II		159.08028	II	II							
PMGC_2574	SSR	NA	NA				105.60371	XV	XI							
PMGC_2140	SSR	159/161	159/159	VII	VII		36.28771	VII	VII							
PMGC_2585	SSR	153/153	155/155				148.90882	XV	XV							
PMGC_2163	SSR	250/274	250/264	X	X		60.65665	X	X							
PMGC_2599	SSR	NA	NA				41.56247	XIII	XIII							
PMGC_2603	SSR	NA	NA				138.13522	VI	VI							
PMGC_2217	SSR	175/175	177/181	VI	VI	VI		no hit	XVIII							
PMGC_2235	SSR	124/124	118/130	IV		IV	88.5733	IV	IV							
PMGC_2270	SSR	120/126	115/115	IV	IV	IV	216.3629	IV	scaffold_163							
PMGC_2289	SSR	64/70	64/64	XIX			123.78194	XIX	XIX							
PMGC_2614	SSR	273/273	273/273		XVI		58.95039	XVI	XVI							
PMGC_2328	SSR	107/109	107/107	VI		VI	221.64299	VI	VI							
PMGC_2660	SSR	111/111	111/111				137.83399	II								
PMGC_2675	SSR	199/AN	195/AN				9.82332	V	V							
PMGC_2385	SSR	163/170	155/181	I	I		460.95691	I	I							
PMGC_2696	SSR	NA	NA		VIII			several hits	several hits							
PMGC_2702	SSR	NA	NA		XIX			several hits	several hits							
PMGC_2419	SSR	106/106	106/113	X			99.95436	X	X							
PMGC_2423	SSR	136/163	163/178	VI	VI	VI	203.49732	VI	VI							
PMGC_2730	SSR	NA	NA				148.57022	VII	VII							
PMGC_2501	SSR	249/255	255/255	III	III		84.33939	III	III							
PMGC_2737	SSR	NA	NA		XII		105.87866	XII	XII							
PMGC_2765	SSR	NA	NA				84.78413	XI	XI							
PMGC_2794	SSR	NA	NA				150.91274	VII	VII							
PMGC_2515	SSR	225/AN	135/224	IV	IV		166.02313	XIV	XIV							
PMGC_2806	SSR	NA	NA				205.2104	X	X							
PMGC_2818	SSR	NA	NA		II		5.41761	II								
PMGC_2826	SSR	233/233	233/233		IV	IV	102.86453	IV	IV							
PMGC_2838	SSR	149/149	149/149			V	169.3698	V	V							
PMGC_2525	SSR	205/210	AN/AN	XVIII	XVIII		90.21159	XVIII	XVIII							
PMGC_2840	SSR	NA	NA				220.33839	V	V							
PMGC_2847	SSR	NA	NA		XIII		118.95024	XIII	Scaffold_41							
PMGC_2536	SSR	219/238	234/242	V			131.84902	V	V							
PMGC_2550	SSR	169/177	171/177	IV ?	I		35.51676	I	I							
PMGC_2858a	SSR	154/154	156/156		III		53.16866	III								
PMGC_2858b	SSR	NT	NT		III		63.64059	III								
PMGC_2556	SSR	132/141	132/141	VI			243.07122	VI	VI							
PMGC_2873	SSR	146/146	146/146		V	V	169.37011	V	V							
PMGC_2573	SSR	124/124	109/124	X	X		119.66406	X	X							
PMGC_2879b	SSR	NT	NT				63.63983	III								
PMGC_2578	SSR	179/179	179/183	VI	VI	VI	84.45322	VI	VI							
PMGC_2598	SSR	131/135	136/141	VIII			26.46641	VIII	VIII							
PMGC_2606	SSR	222/240	236/245	V	V		131.85066	V	V							
PMGC_325	SSR	NA	NA		XVII		131.4913	XVII	XVII							
PMGC_2607	SSR	185/188	184/218	VIII	VIII	VIII	30.14917	VIII	VIII							
PMGC_409	SSR	NA	NA		VIII	VIII	87.71803	VIII	VIII							
PMGC_420	SSR	NA	NA			XIV	72.75083	XIV	XIV							
PMGC_2610	SSR	103/105	103/105	VIII	VIII		64.97042	VIII	VIII							
PMGC_2611	SSR	272/278	278/278	III	III		84.33886	III	III							
PMGC_2658	SSR	274/274	265/274	XIII	XIII		40.949	XIII	XIII							
PMGC_2679	SSR	131/133	131/131	XV	XV		32.65339	XV	XV							
PMGC_2709	SSR	219/238	221/223	II	II		120.112	II	II							
PMGC_2718	SSR	125/125	123/125	VI	VI		98.05469	VI	VI							
PMGC_571	SSR	NA	NA				99.33298	XIV	XIV							
PMGC_575	SSR	NA	NA			VII	327.0527	I								
PMGC_576	SSR	NA	NA		V		131.84902	V	V							
PMGC_2731	SSR	232/235	230/232	I	I		317.52519	I	I							
PMGC_2803	SSR	193/198	193/198	XVII	XVII		26.21937	IX	XVII							
PMGC_639	SSR	117/117	117/117		V		17.42715	V	V							
PMGC_648	SSR	NA	NA		XVII	L	18.44971	XVII	XVII							
PMGC_649	SSR	129/133	129/133				90.90554	XIII	XIII							
PMGC_2804	SSR	77/93	99/99	XVI			15.74601	XVI	XVI							
PMGC_683	SSR	NA	NA				138.00619	II								
PMGC_684	SSR	NA	NA				70.66536	II	II							
PMGC_690	SSR	NA	NA				48.75451	XV	XV							
PMGC_93	SSR	NA	NA		I	I	206.34129	I	no hit							
PTAG1	SSR	NA	NA					several hits	several hits							
PMGC_2832	SSR	202/206	202/206	XVII			26.21937	IX	XVII							
PMGC_2839	SSR	223/223	217/225	V	V	V	232.54482	V	V							
Ptr03	SSR	NA	NA					several hits	several hits							
Ptr04	SSR	227/227	227/227				43.8738	III								
Ptr05	SSR	NA	NA					several hits	several hits							
Ptr06	SSR	215/215	215/215				60.29627	I								
PMGC_2852	SSR	99/101	103/142	I	I	I	145.48351	I	I							
Ptr08	SSR	158/158	158/158				93.36749	VIII	VIII							
Ptr11	SSR	165/165	165/165				149.67746	XV								
PMGC_2855	SSR	152/152	138/152	X	X		119.66383	X	X							
Ptr14	SSR	219/219	219/219				152.84685	III								
Ptr15	SSR	NA	NA					several hits	several hits							
PMGC_2861	SSR	238/AN	240/240	XI			12.40245	IX	scaffold_170							
PMGC_2866	SSR	203/215	213/215	XI	XI	XI	153.64686	XI	XI							
PMGC_2879a	SSR	180/182	186/180	III	III	III	53.1697	III	III							
PMGC_2881	SSR	162/180	162/169	IV		IV	141.71603	IV	IV							
PMGC_2885	SSR	335/335	337/339	XII	XII		107.78032	XII	XII							
PMGC_2889	SSR	222/228	222/224	XVII	XVII		66.51335	XVII	XVII							
PMGC_333	SSR	94/94	101/107	XI	XI	XI	173.61598	XI	XI							
TGRC_AG2	SSR	NA	NA				32.07937	XI	XI							
PMGC_422	SSR	294/298	292/294	II			48.41308	II	II							
WPMS_10	SSR	NA	NA		III	III	187.67111	III								
PMGC_433	SSR	198/204	196/204	XVI		XVI	106.9012	XIX	scaffold_125							
PMGC_451	SSR	209/AN	209/241	XVII	XVII	C	43.34513	XVII	scaffold_21039							
PMGC_456	SSR	100/102	102/102	II			137.83391	II	II							
PMGC_486	SSR	144/148	136/146	III	III	III	195.49796	III	III							

PMGC_510	SSR	162/162	162/168	X	X		99.95391	X	X							
PMGC_520	SSR	141/141	126/141	XV	XV		117.98347	XV	XV							
PMGC_562	SSR	267/269	267/267	VII	VII		5.77466	VII	scaffold_12371							
WPMS_18	SSR	NA	NA		I		6.62094	I								
WPMS_19	SSR	NA	NA		V		200.48143	V	V							
WPMS_2	SSR	NA	NA		XVI			several hits	several hits							
PMGC_607	SSR	164/175	169/177	VII	VII	VII	9.73782	VII	VII							
PMGC_61	SSR	376/376	369/376	VIII	VIII	VIII	64.35292	VIII	VIII							
PMGC_667	SSR	192/213	183/192	II	II	II	35.64417	II	II							
Ptr01	SSR	280/280	274/280	IV	IV		64.11786	IV	IV							
Ptr02	SSR	235/238	235/238	IX			82.90993	IX	IX							
Ptr07	SSR	275/275	261/273	XII	XII		6.38576	XII	XII							
Ptr12	SSR	270/270	270/284	VII			28.87046	VII	no hit							
GCPM_1061-1	SSR	188/188	188/188				383.51827	I								
GCPM_1274	SSR	179/215	179/215				457.29846	I								
GCPM_217-1	SSR	136/136	136/136				405.62652	I								
GCPM_2240-1	SSR	240/240	240/240				394.58912	I								
GCPM_2302-1	SSR	230/230	230/230				372.84032	I								
GCPM_2952-1	SSR	167/173	167/167				374.20341	I								
GCPM_2966-1	SSR	205/205	205/205				430.60961	I								
GCPM_3451-1	SSR	135/135	127/184				410.90965	I								
GCPM_3485-1	SSR	225/225	225/225				364.14294	I								
GCPM_669-2	SSR	252/252	252/252				388.40013	I								
GCPM_862-1	SSR	155/155	155/155				358.58272	I								
GCPM_875-1	SSR	156/156	156/156				365.2304	I								
WPMS_1	SSR	160/176	176/182	XIII	XIII		71.17779	XIII	no hit							
WPMS_11	SSR	240/246	216/226	II	II		119.50067	II	II							
WPMS_12	SSR	187/191	184/187	VI	VI	VI	194.71767	VI	VI							
WPMS_13	SSR	141/143	159/165	VIII	VIII	VIII	112.33658	VIII	VIII							
WPMS_14	SSR	273/273	290/304	V	V		200.48185	V	II							
WPMS_15	SSR	220/239	236/239	V			220.33839	V	V							
WPMS_16	SSR	170/176	170/176	VII		VII	149.62206	VII	VII							
WPMS_17	SSR	154/154	154/167	VII	VII	VII	86.96109	VII	scaffold_181							
WPMS_20	SSR	245/251	257/263	XIII	XIII		154.29213	XIII	XIII							
WPMS_21	SSR	310/310	270/346	II			50.57778	II	II							
WPMS_22	SSR	141/144	110/135	XIII	XIII		29.38283	XIII	XIII							
WPMS_3	SSR	298/313	302/315	XII	XII		37.57776	XII	XII							
WPMS_4	SSR	295/295	272/281	VI	VI			no hit	VI							
WPMS_5	SSR	278/AN	267/280	XII	XII		92.08512	XII	XII							
WPMS_6	SSR	228/236	230/236	XVI	XVI	XII	101.76967	XVI	no hit							
WPMS_7	SSR	261/286	257/286	VI	VI		218.87093	VI	no hit							
WPMS_8	SSR	257/270	257/AN	XVII				no hit	no hit							
WPMS_9	SSR	287/287	279/287	VI		VI	236.13768	VI	VI							
LG_IV_11_195	SNP	A	R	IV			103.80129	IV	IV	GATAATGCAC[T/C]AAC	eugene3.00	POPTR_000	AT5G15640	mitochon	similar to	similar to
LG_IV_18_239P	SNP	C	S	IV			169.75396	IV	IV	ATATTAGAAT[G/C]GAT	gw1.IV.639.	POPTR_000	AT4G15850/A	ATRH1;	no poplar v2	similar to
LG_IV_18_351P	SNP	M	C	IV			169.75396	IV	IV	ATCTTGATGA[A/C]GAA						similar to
LG_IV_20_227	SNP	C	S	IV			170.8127	IV	IV	TTAATTCCTC[G/C]CAA	estExt_fgen	POPTR_000	AT4G36360	BGAL3	similar to beta-	similar to beta-
LG_IV_21_405P	SNP	C	Y	IV			7.95646	IV	IV	AGAGGCTTGC[T/C]GAG	gw1.IV.1616	POPTR_000	AT1G11720	ATSS3	similar to starch	similar to
LG_IV_21_42P	SNP	R	G	IV			7.95646	IV	IV	GTTCAATTGC[T/C]AGTT						similar to
LG_IV_31_46	SNP	T	Y	IV			65.17612	IV	IV	GGATTTGGGC[C/T]TCA	gw1.IV.3768	POPTR_000	AT5G17240	SDG40	no poplar v2	KOG:1337N-
LG_IV_35_438	SNP	C	Y	IV			176.11243	IV	IV	TCAAAATGAA[A/G]AT	gw1.IV.950.	POPTR_000	AT4G34830		no poplar v2	Panther:10483
LG_IX_10_150P	SNP	C	Y	X			196.02045	X	IX	ACGGAATAGG[C/T]CTC	eugene3.00	POPTR_000	AT2G16270	unknown	no poplar v2	para-
LG_IX_10_8P	SNP	M	A	X			196.02045	X	IX	AAGATTACCA[A/C]TTA						para-
LG_IX_20_46	SNP	G	K	IX			94.0195	IX	IX	CAAGATTGCT[G/T]CGA	gw1.IX.1988	POPTR_000	AT2G23460	XLG1	no poplar v2	Panther:10218
LG_IX_23_316	SNP	C	S	IX			64.45417	IX	IX	CAAAACTGAT[C/G]AA	gw1.IX.3264	POPTR_000	AT3G08505	zinc	no poplar v2	Pfam:00097 Zi
LG_IX_27_110	SNP	G	R	IX			33.58643	IX	IX	TTACCGGTAC[C/T]CAA	gw1.IX.4548	POPTR_000	AT1G14370	APK2A	no poplar v2	Panther:23258
LG_IX_29_150P	SNP	M	C	IX			13.80129	IX	IX	TTGAATCCTC[C/A]AAC	grail3.00011	POPTR_000	AT2G28110	FRA8	exostosin family	exostosin
LG_IX_29_230P	SNP	A	W	IX			13.80129	IX	IX	TTATTTATGT[A/T]TTTT						exostosin
LG_IX_4_157	SNP	T	Y	IX			107.58261	IX	IX	TAACAAAAGA[G/A]TA	grail3.00010	POPTR_000	AT4G38640	choline	similar to	similar to
LG_IX_7_149P	SNP	R	G	X			138.90989	X	IX	ATACAAGTGA[C/T]ATT	fgenesh4_p	POPTR_000	AT4G35070		no poplar v2	similar to
LG_IX_7_209P	SNP	A	W	X			138.90989	X	IX	CAACTATTTT[A/T]TGTT						similar to
LG_IX_8_254	SNP	G	R	X			153.79291	X	IX	AAAATGTCCA[T/C]TCT	estExt_fgen	POPTR_000	AT4G34540	isoflavon	isoflavone	similar to
LG_V_11_578P	SNP	G	R	V			172.85301	V	V	GCAGGCGGGT[A/G]GA	gw1.V.4245.	POPTR_000	AT5G67360	ARA12;	no poplar v2	Panther:10795
LG_V_11_89P	SNP	R	A	V			172.85301	V	V	CATAAACAAA[T/C]GGC						Panther:10795
LG_V_16_88	SNP	G	R	V			51.03027	V	V	AAAAGTGATT[G/A]CTC	gw1.V.2759.	POPTR_000	AT5G10320	unknown	no poplar v2	no annotation
LG_V_29_284P	SNP	W	A	V			60.12743	V	V	ACAAAACCGT[T/A]TAG	gw1.V.3090.	POPTR_000	AT5G09970	CYP78A7;	no poplar v2	Panther:19383
LG_V_29_374P	SNP	A	M	V			60.12743	V	V	CGACATGGAG[A/C]AC						Panther:19383
LG_V_37_193	SNP	K	G	V			158.38372	V	V	TAGGAGCGAA[C/A]CA	eugene3.00	POPTR_000	AT1G09730	Ulp1	Ulp1 protease	Panther:12606
LG_V_40_258	SNP	A	M	V			207.48948	V	V	CTCAGGTAGT[A/C]ACC	estExt_fgen	POPTR_000	AT1G43170	ARP1	similar to	similar to
LG_V_5_340	SNP	W	T	V			177.86055	V	V	ATAATCTTAC[T/A]TAC	estExt_fgen	POPTR_000	AT1G07670	calcium-	no poplar v2	Panther:11939
LG_V_9_46	SNP	R	A	V			194.29066	V	V	AAAAGAGGTG[T/C]CAT	estExt_Gen	POPTR_000	AT1G44110	CYCA1;1	similar to cyclin;	similar to
LG_VI_1_417	SNP	A	R	VI			63.91818	VI	VI	ACCCATAAGA[G/A]CA	estExt_fgen	POPTR_000	AT5G02230	haloacid	no poplar v2	Panther:18901
LG_VI_15_231	SNP	C	Y	VI			196.11247	VI	VI	CACTGGGCCC[C/T]GAT	estExt_fgen	POPTR_000	AT1G20970		no poplar v2	no annotation
LG_VI_5_323	SNP	G	K	VI			223.54632	VI	VI	TTTTTATTAC[G/T]GCAT	eugene3.00	POPTR_000	AT5G21940	unknown	similar to MTD1.	similar to
LG_VII_1_171	SNP	G	K	VII			24.0225	VII	VII	AAGGATGAGA[G/T]TA	estExt_Gen	POPTR_000	AT5G09760	pectinest	pectinesterase	pectinesterase
LG_VII_11_203	SNP	Y	T	VII			139.90204	VII	VII	AATCTTACTG[G/A]CTG	estExt_Gen	POPTR_000	AT5G45600	GAS41;	no poplar v2	Panther:23195
LG_VII_20_262	SNP	S	C	VII			15.28374	VII	VII	CCATTACAGT[G/C]GGG	eugene3.00	POPTR_000	AT1G05270	TraB	TraB family	TraB family
LG_VII_32_167	SNP	G	R	VII			116.46597	VII	VII	CAAGATAAAA[C/T]GCC	gw1.VII.304	POPTR_000	AT3G50590/A	nucleotid	no poplar v2	transducin
LG_VII_8_128	SNP	Y	C	VII			133.26978	VII	VII	AGGAATAGCG[G/A]CT	eugene3.00	POPTR_000	AT4G34910	DEAD/DE	no poplar v2	Panther:10967
LG_VIII_1_153	SNP	R	A	VIII			33.13151	VIII	VIII	CATTTTGCAT[G/A]ACC	estExt_fgen	POPTR_000	AT2G39700	ATEXPA4	similar to alpha-	similar to
LG_VIII_10_96	SNP	G	K	VIII			99.32033	VIII	VIII	GGAGAGCCTT[G/T]TCC	estExt_Gen	POPTR_000	AT5G19300		similar to	similar to
LG_VIII_33_250P	SNP	T	Y	VIII			56.79638	VIII	VIII	GTAATTGTTT[T/C]ATTA	estExt_fgen	POPTR_000	AT1G69400	transduci	transducin	transducin
LG_VIII_33_288P	SNP	R	G	VIII			56.79638	VIII	VIII	TAGTTCTTAT[A/G]GTAT						transducin
LG_VIII_34_138	SNP	A	W	VIII			61.78556	VIII	VIII	GAGACACAAT[T/A]TGA	fgenesh4_p	POPTR_000	AT3G62270	anion	no poplar v2	Panther:11453
LG_VIII_36_86	SNP	M	C	VIII			72.89749	VIII	VIII	CACCGTCGAA[C/A]CTC	eugene3.00	POPTR_000	AT3G25700	chloropla	no poplar v2	Panther:13683
LG_VIII_9_534	SNP	G	S	VIII			98.73826	VIII	VIII	CTGGCTAGGT[G/C]ACT	gw1.VIII.10	POPTR_000	AT5G22000	RHF2A	no poplar v2	Pfam:00097 Zi
LG_X_31_141	SNP	C	M	X			103.87724	X	X	CAAAGCTTTT[C/A]AAT	gw1.X.5864.	POPTR_001	AT3G22520	unknown	no poplar v2	no annotation
LG_X_39_332	SNP	T	Y	X			171.37756	X	X	TATAATATCA[G/A]CGC	gw1.X.1601.	POPTR_001	AT2G40320	unknown	no poplar v2	Pfam:03005 A
LG_X_43_252	SNP	R	A	X			201.21139	X	X	GTTTTTCTCC[G/A]AAGT	fgenesh4_p	POPTR_001			similar to	GO:0045132me
LG_X_5_307	SNP	K	G	X			165.71177	X	X	GATCAGACTT[C/A]ATA	grail3.00220	POPTR_001	AT1G79350	EMB1135	no poplar v2	KOG:1513Nucl
LG_X_6_120	SNP	C	Y	III			151.91006	III	X	CTTCAGTCTG[C/T]CTGT	estExt_Gen	POPTR_001	AT3G06860	MFP2	no poplar v2	similar to
LG_X_PopCHT1_2_470	SNP	Y	T	X			144.98273	X	X	AGTTTAATAA[G/A]TTC						Panther:23309
LG_XI_6_226	SNP	K	G	XI			134.27572	XI	XI	GAAATTGTTA[G/T]CAT	gw1.XI.3477	POPTR_001	AT1G16970	KU70	similar to Ku70-	KOG:2327DNA-
LG_XI_7_114	SNP	A	M	XI			137.57299	XI	XI	TTTTAGTATG[C/A]TACC	eugene3.00	POPTR_001	AT1G78810	unknown	similar to	Pfam:03470 X
LG_XI_9_93	SNP	Y	C	XI			139.6145	XI	XI	AGGATTCCCC[G/A]GGT	estExt_Gen	POPTR_001	AT2G42590	GRF9	similar to 14-3-3	similar to 14-3-
LG_XII_16_194	SNP	R	G	XII			132.83263	XII	XII	GCGGACCTGC[G/A]TCC	estExt_fgen	POPTR_001	AT4G25620	hydroxyp	hydroxyproline-	hydroxyprolin
LG_XII_4_448	SNP	T	K	XII			96.63805	XII	XII	AACAAAACCT[C/A]CA	eugene3.00	POPTR_001	AT5G63110	HDA6	similar to	similar to
LG_XII_8_202	SNP	C	Y	XII			104.41907	XII	XII	CTTTCGTTAG[C/T]TTTT	gw1.XII.125	POPTR_001	AT3G48440	zinc	no poplar v2	Panther:12547
LG_XIV_10_121	SNP	C	Y	XIV			47.9603	XIV	XIV	TAATCAAATA[C/T]TAA	gw1.XIV.72	POPTR_001	AT3G60740	TTN1	similar to	similar to
LG_XIV_12_504	SNP	C	M	XIV			48.49707	XIV	XIV	TTAAGAGTTT[C/A]AAC	estExt_Gen	POPTR_001	AT3G60750	transketo	no poplar v2	Pfam:00456 T

LG_XIV_13_535P	SNP	K	G	II			171.32549	II	XIV	TTTGTACCAT[G/T]ACTA	gw1.XIV.27	POPTR_000	AT3G62770	AtATG18a	no poplar v2	Panther:11227
LG_XIV_13_560P	SNP	T	Y	II			171.32549	II	XIV	ATGATATCAC[G/A]AAT						Panther:11227
LG_XIV_4_130	SNP	C	Y	XIV			133.68146	XIV	XIV	AATGTTTTCA[C/T]CTTA	estExt_Gen	POPTR_001	AT5G39740	60S	no poplar v2	Panther:23410
LG_XIV_8_220	SNP	Y	C	XIV			47.29837	XIV	XIV	TTTCATGATT[C/T]CTTA	grail3.00350	POPTR_001	AT2G45190	AFO	similar to	similar to
LG_XIX_12_165	SNP	C	M	XIX			148.58134	XIX	XIX	CAACTTGAGA[C/A]GTA	estExt_fgen	POPTR_001	AT4G03280	PETC	similar to	similar to
LG_XIX_19_314	SNP	G	S	XIX			87.00672	XIX	XIX	ATTTCTATGG[G/C]AAG	eugene3.00	POPTR_001	AT3G03050	CSLD3	similar to	similar to
LG_XIX_3247-	SNP	G	R	XIX			78.8561	XIX	XIX	AGTAACCAAT[C/T]TGT	estExt_fgen					Pfam:05770 I
LG_XIX_39_231P	SNP	M	A	XIX			129.19031	XIX	XIX	CTTAAGCAAG[T/G]TAT	gw1.XIX.219	POPTR_001	AT1G34270	exostosin	exostosin family	exostosin
LG_XIX_39_303P	SNP	C	M	XIX			129.19031	XIX	XIX	GTATTTGTTT[A/C]AGTT						exostosin
LG_XIX_4218_4220nco	SNP	T	W	XIX			88.18298	XIX	XIX	TGTGCTGGGA[T/A]CAA	eugene3.00	POPTR_001	AT4G24340	phosphor	no poplar v2	Panther:21234
LG_XIX_5017-	SNP	Y	C	XIX			51.95081	XIX	XIX	CTGGTCCGTT[G/A]CTTC	eugene3.00	POPTR_001	AT3G13460	ECT2;	no poplar v2	Panther:12357
LG_XIX_5302_5303_46	SNP	C	Y	XIX			54.87637	XIX	XIX	GGTCGGCTCA[C/T]AAC	fgenesh4_p	POPTR_001	AT4G38040	exostosin	no poplar v2	KOG:1021Acety
LG_XIX_5603_5606_17	SNP	C	M	XIX			58.74007	XIX	XIX	GTGTAAACGT[G/T]AAT	fgenesh4_p	POPTR_001	AT5G15120	unknown	no poplar v2	GO:0047800cys
LG_XIX_5787_5789_78	SNP	G	R	XIX			60.57567	XIX	XIX	ACTATTTTGA[C/T]AGA	fgenesh4_p	POPTR_001	AT3G04440	unknown	similar to	KOG:1362Choli
LG_XV_16_285	SNP	C	Y	XV			44.9811	XV	XV	CCAGTTACCA[G/A]GAA	gw1.XV.966	POPTR_001	AT3G18040	MPK9	similar to MAP	similar to MAP
LG_XV_5_58	SNP	M	A	XV			135.10933	XV	XV	TCTATACAGA[G/T]TGA	gw1.XV.320	POPTR_001	AT5G52110	HCF208	no poplar v2	Pfam:11152 P
LG_XVI_15_531P	SNP	G	R	XVI			133.20795	XVI	XVI	TGAACCAAAC[G/A]AA	gw1.XVI.23	POPTR_001	AT2G38370	unknown	no poplar v2	Pfam:05701 P
LG_XVI_15_91P	SNP	K	G	XVI			133.20795	XVI	XVI	TTTCTAGTAT[C/A]TGCA						Pfam:05701 P
LG_XVI_PopLTP1_137	SNP	R	G	XVI			83.1521	XVI	XVI	TGGGCCAGAA[G/A]CC						protease
LG_XVI_PopLTP1_368	SNP	C	Y	XVI			83.1521	XVI	XVI	GTAAGGTAAA[C/T]ATT						protease
LG_XVII_13_97	SNP	K	G	XVII			125.95087	XVII	XVII	ATGGTTCTTT[G/T]GTAA	eugene3.00	POPTR_001	AT1G13960/A	WRKY4;	no poplar v2	Pfam:03106
LG_XVII_16_207	SNP	C	M	XVII			132.23161	XVII	XVII	GGTCAGAAAC[C/A]AT	fgenesh4_p	POPTR_001	AT1G74160	unknown	no poplar v2	Panther:21726
LG_XVII_6_236	SNP	C	S	XVII			105.33315	XVII	XVII	ATTGAGTTTT[G/C]TTTC	estExt_fgen	POPTR_001	AT3G55250	unknown	no poplar v2	no annotation
LG_XVII_7_247	SNP	W	A	XVII			106.85635	XVII	XVII	CAAAGGTGAG[T/A]ATC	gw1.XVII.76	POPTR_001	AT3G28345	ABC	ABC transporter	ABC
LG_XVII_8_153	SNP	C	Y	XVII			113.81019	XVII	XVII	AGTAGCAGGT[G/A]CTT	estExt_fgen	POPTR_001	AT1G74690	IQD31 (IQ-	calmodulin-	calmodulin-
LG_XVII_9_183	SNP	C	M	XVII			115.03635	XVII	XVII	AATGTTTTAA[G/T]AAA	estExt_Gen	POPTR_001	AT4G14960	TUA6;	no poplar v2	Pfam:00091 T
LG_XVIII_11_214	SNP	W	T	XVIII			133.46458	XVIII	XVIII	GAGCAGAAAA[A/T]TA	gw1.XVIII.1	POPTR_001	AT2G42160	zinc	no poplar v2	Panther:10420
LG_XVIII_13_298	SNP	T	W	XVIII			142.77391	XVIII	XVIII	GTGGCAACTT[A/T]AAA	estExt_fgen	POPTR_001	AT4G30020	subtilase	no poplar v2	Panther:10795
LG_XVIII_16_138P	SNP	A	R	XVIII			45.08968	XVIII	XVIII	TCAAAACCCA[C/T]CGG	fgenesh4_p	POPTR_001	AT4G32980	ATH1	no poplar v2	KOG:0773Trans
LG_XVIII_16_491P	SNP	K	T	XVIII			45.08968	XVIII	XVIII	TGAATAAAGC[A/C]ATT						KOG:0773Trans
scaffold_121_1_246	SNP	R	A	III			4.35036	scaffold_22	scaffold_121_1	AGAACTCTCC[C/T]ATTT	fgenesh4_p	POPTR_000	AT4G25880	APUM6	no poplar v2	KOG:1488Trans
scaffold_123_1_361	SNP	G	R	XVII			122.30232	V	scaffold_123_1	ACAGTGCTAA[C/T]ATC	eugene3.01	POPTR_000	AT1G74640	unknown	similar to	Panther:10992:
scaffold_140_1_376	SNP	Y	C	XI			186.4902	XI	scaffold_140_1	AAGATGACAG[G/A]AT	eugene3.01	POPTR_001	AT1G55740	AtSIP1	similar to	similar to
scaffold_142_1_239	SNP	K	T	XIII			137.79205	XIII	scaffold_142_1	AGGATCCTTC[T/G]TCA	grail3.01420	POPTR_001	AT5G39740	60S	no poplar v2	Panther:23410
scaffold_142_2_383	SNP	G	S	XIII			139.31747	XIII	scaffold_142_2	AATCTGTTCT[G/C]GCC	estExt_fgen	POPTR_001	AT4G03560	ATTPC1	similar to	similar to
scaffold_29_5_71	SNP	C	S	I			24.03669	I	scaffold_29_5	ACTATTTCTT[G/C]TCAC	eugene3.00	POPTR_000	AT4G18210	ATPUP10;	no poplar v2	Pfam:03151 T
scaffold_40_1_294	SNP	T	Y	XIV			28.63739	XIV	scaffold_40_1	TGATTTTATA[C/T]GAA	estExt_fgen	POPTR_001	AT1G47670	amino	amino acid	amino acid
scaffold_41_3_134	SNP	C	S	XIII			99.88449	XIII	scaffold_41_3	CAAGCAGCCT[G/C]TGT	eugene3.00	POPTR_001	AT1G58350	ZW18	similar to	similar to
scaffold_86_1_198	SNP	T	Y	IV			231.10246	IV	scaffold_86_1	AGTAGCAGGT[G/A]CTT	estExt_fgen	POPTR_000	AT4G19190	zinc	zinc knuckle	zinc knuckle
PM01	PAMSA	211/211	211/211				251.68997	I	I	**F1***CACGACGTTGTAAAA*						
PM02	PAMSA	218/218	218/218				246.41675	I	I	**F1***CACGACGTTGTAAAA*						
PM05	PAMSA	197/197	197/197				39.53603	IX	I	**F1***CACGACGTTGTAAAA*						
PM06	PAMSA	172/172	172/172					no hit	no hit	**F1***CACGACGTTGTAAAA*						
PM13	PAMSA	172/172	172/172				109.43553	III	III	**F1***CACGACGTTGTAAAA*						
PM14	PAMSA	268/268	268/268				109.43597	III	III	**F1***CACGACGTTGTAAAA*						
PM03	PAMSA	199/202	199/199	I			279.27084	I	I	**F1***CACGACGTTGTAAAA*						
PM04	PAMSA	132/132	132/135	I			279.27122	I	I	**F1***CACGACGTTGTAAAA*						
PM07	PAMSA	193/193	193/196	II			80.13486	II	II	**F1***CACGACGTTGTAAAA*						
PM08	PAMSA	256/259	256/256	II			83.26561	II	II	**F1***CACGACGTTGTAAAA*						
PM09	PAMSA	255/258	255/255	II			125.45144	II	II	**F1***CACGACGTTGTAAAA*						
PM10	PAMSA	231/231	231/234	II			125.45167	II	II	**F1***CACGACGTTGTAAAA*						
PM11	PAMSA	234/234	234/237	II			133.57563	II	II	**F1***CACGACGTTGTAAAA*						
PM12	PAMSA	231/234	231/231	II			134.41927	II	II	**F1***CACGACGTTGTAAAA*						
PM15	PAMSA	266/269	266/266	X			70.60084	X	III	**F1***CACGACGTTGTAAAA*						
EacgMcaa146	AFLP*	146/146	AN/AN	I												
EacgMcaa201	AFLP	201/201	AN/AN	II												
EacgMcaa378	AFLP	AN/AN	378/378	VI												
EacgMcac122	AFLP	122/122	AN/AN	VII												
EacgMcac192	AFLP	192/192	AN/AN	I												
EacgMcac244	AFLP	AN/AN	244/244	I												
EacgMcat100	AFLP	100/100	AN/AN	X												
EacgMcat147	AFLP	147/147	AN/AN	XV												
EacgMcat176	AFLP	AN/AN	176/176	XIV												
EacgMcat219	AFLP	219/219	AN/AN	XVII												
EacgMcat222	AFLP	222/222	AN/AN	I												
EacgMcat251	AFLP	AN/AN	251/251	XVII												
EacgMcat259	AFLP	AN/AN	259/259	XVII												
EacgMcat282	AFLP	AN/AN	282/282	IV												
EacgMcat358	AFLP	AN/AN	358/358	I												
EacgMcat404	AFLP	404/404	AN/AN	XII												
EacgMcga123	AFLP	123/123	AN/AN	VIII												
EacgMcga197	AFLP	197/197	AN/AN	VIII												
EacgMcga258	AFLP	AN/AN	258/258	X												
EacgMcga343	AFLP	AN/AN	343/343	XIV												
EacgMctc115	AFLP	115/115	AN/AN	V												
EacgMctc178	AFLP	AN/AN	178/178	XIV												
EacgMctc202	AFLP	202/202	AN/AN	I												
EacgMctc281	AFLP	AN/AN	281/281	V												
EacgMctg107	AFLP	AN/AN	107/107	XVI												
EacgMctg116	AFLP	116/116	AN/AN	I												
EacgMctg180	AFLP	AN/AN	180/180	XIII												
EacgMctg195	AFLP	195/195	AN/AN	XII												
EacgMctg223	AFLP	AN/AN	223/223	IV												
EacgMctg260	AFLP	AN/AN	260/260	VII												
EaggMcag102	AFLP	AN/AN	102/102	XV												
EaggMcag129	AFLP	AN/AN	129/129	XVII												
EaggMcag149	AFLP	149/149	AN/AN	XI												
EaggMcag295	AFLP	AN/AN	295/295	VI												
EaggMcag508	AFLP	508/508	AN/AN	XIII												
EaggMcag61	AFLP	AN/AN	61/61													
EaggMctt144	AFLP	AN/AN	144/144	IX												
EaggMctt221	AFLP	221/221	AN/AN	XIX												
EaggMctt240	AFLP	AN/AN	240/240	I												
EaggMctt400	AFLP	AN/AN	400/400	I												
EaggMctt68	AFLP	AN/AN	68/68	VIII												

Legend :

*41 AFLP markers presented in this table are polymorphic, we tested 101

AN (null allele)

NA (Not Amplified)

NT (not tested) because the same marker has two matches

Bibliographie

Abraham P, Adams R, Giannone RJ, Kalluri U, Ranjan P, Erickson B, Shah M, Tuskan GA, Hettich RL (2012) Defining the Boundaries and Characterizing the Landscape of Functional Genome Expression in Vascular Tissues of Populus using Shotgun Proteomics. J Proteome Res 11 (1):449-460. doi:10.1021/pr200851y

Abrahamson LP, Volk TA, Kopp RF, White EH, Ballard JL (2002) Willow Biomass Producer's Handbook. Syracuse State University of New York, College of Environmental Science and Forestry:31

ADEME (2010) Analyses de Cycle de Vie appliquées aux biocarburants de première génération consommés en France. 236 pp.

ADEME, AFGNV, ATEE-Club-Biogaz, GDF-SUEZ, IFP, MEEDDAT (2008) État des lieux et potentiel du biométhane carburant. Synthèse 8pp.

Al-Afas N, Marron N, Ceulemans R (2007) Variability in Populus leaf anatomy and morphology in relation to canopy position, biomass production, and varietal taxon. Annals of Forest Science 64 (5):521-532. doi:10.1051/forest:2007029

Al Afas N, Marron N, Van Dongen S, Laureysens I, Ceulemans R (2008a) Dynamics of biomass production in a poplar coppice culture over three rotations (11 years). For Ecol Manage 255 (5-6):1883-1891. doi:10.1016/j.foreco.2007.12.010

Al Afas N, Marron N, Zavalloni C, Ceulernansk R (2008b) Growth and production of a short-rotation coppice culture of poplar-IV: Fine root characteristics of five poplar clones. Biomass Bioenerg 32 (6):494-502. doi:10.1016/j.biombioe.2007.11.007

Alberto F (2010) Différentiation phénologique et moléculaire du chêne sessile le long de gradients environnementaux. Université de Bordeaux, thèse de doctorat, 207 pp.

Alker G, Bruton C, Richards K (2005) Full-scale implementation of SRC-systems: Assessment of Technical and Non-Technical Barriers. IEA Bioenergy Task 30

Alvarez ME, Nota F, Cambiagno DA (2010) Epigenetic control of plant immunity. Mol Plant Pathol 11 (4):563-576. doi:10.1111/j.1364-3703.2010.00621.x

Andersson-Gunneras S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, Nilsson P, Henrissat B, Moritz T, Sundberg B (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. doi:10.1111/j.1365-313X.2005.02584.x

Aronsson P, Dimitriou I, Rosenqvist H (2006) Solutions for the safe application of wastewater and sludge for high efficient biomass production in Short-Rotation-Plantations. D4 – Report on ongoing research and gaps in SRP knowledge.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465 (7298):627-631. doi:10.1038/nature08800

Aylott MJ, Casella E, Tubby I, Street NR, Smith P, Taylor G (2008) Yield and spatial supply of bioenergy poplar and willow short-rotation coppice in the UK. New Phytologist 178 (2):358-370. doi:10.1111/j.1469-8137.2008.02396.x

Azaiez A, Boyle B, Levee V, Seguin A (2009) Transcriptome profiling in hybrid poplar following interactions with Melampsora rust fungi. Molecular Plant-Microbe Interactions 22 (2):190-200. doi:10.1094/mpmi-22-2-0190

Babiychuk E, Bouvier-Nave P, Compagnon V, Suzuki M, Muranaka T, Van Montagu M, Kushnir S, Schaller H (2008) Allelic mutant series reveal distinct functions for Arabidopsis cycloartenol synthase 1 in cell viability and plastid biogenesis. Proceedings of the National Academy of Sciences of the United States of America 105 (8):3163-3168. doi:10.1073/pnas.0712190105

Babst BA, Sjodin A, Jansson S, Orians CM (2009) Local and systemic transcriptome responses to herbivory and jasmonic acid in Populus. Tree Genetics & Genomes 5 (3):459-474. doi:10.1007/s11295-009-0200-6

Bailleres H, Davrieus F, Pichavant FH (2002) Near infrared analysis as a tool for rapid screening of some major wood characteristics in a eucalyptus breeding program. Annals of Forest Science 59 (5-6):479-490. doi:10.1051/forest:20032

Ball J, Carle J, Del Lungo A (2005) Contribution of poplars and willows to sustainable forestry and rural development. FAO, Unasylva 221 56

Barakat A, Wall PK, DiLoreto S, Pamphilis CWd, Carlson JE (2007) Conservation and divergence of microRNAs in Populus. BMC Genomics 8. doi:doi:10.1186/1471-2164-8-481

Barcaccia G, Meneghetti S, Albertini E, Triest L, Lucchin M (2003) Linkage mapping in tetraploid willows: segregation of molecular markers and estimation of linkage phases support an allotetraploid structure for Salix alba x Salix fragilis interspecific hybrids. Heredity 90 (2):169-180. doi:10.1038/sj.hdy.6800213

Bauer S, Grossmann S, Vingron M, Robinson PN (2008) Ontologizer 2.0 - a multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics 24 (14):1650-1651. doi:10.1093/bioinformatics/btn250

Benomar L, DesRochers A, Larocque GR (2012) The effects of spacing on growth, morphology and biomass production and allocation in two hybrid poplar clones growing in the boreal region of Canada. Trees: Structure and Function 26 (3):939-949. doi:10.1007/s00468-011-0671-6

Bergante S, Facciotto G, Minotta G (2010) Identification of the main site factors and management intensity affecting the establishment of short-rotation-coppices (SRC) in Northern Italy through stepwise regression analysis. Central European Journal of Biology 5 (4):522-530. doi:10.2478/s11535-010-0028-y

Berlin S, Fogelqvist J, Lascoux M, Lagercrantz U, Ronnberg-Wastljung AC (2011) Polymorphism and Divergence in Two Willow Species, Salix viminalis L. and Salix schwerinii E. Wolf. G3 (Bethesda, Md) 1 (5):387-400. doi:10.1534/g3.111.000539

Berlin S, Lagercrantz U, von Arnold S, Ost T, Ronnberg-Wastljung AC (2010) High-density linkage mapping and evolution of paralogs and orthologs in Salix and Populus. Bmc Genomics 11:14. doi:10.1186/1471-2164-11-129

Bernhardsson C, Ingvarsson PK (2011) Molecular Population Genetics of Elicitor-Induced Resistance Genes in European Aspen (Populus tremula L., Salicaceae). PLoS ONE 6 (9). doi:10.1371/journal.pone.0024867

Bernhardsson C, Ingvarsson PK (2012) Geographical structure and adaptive population differentiation in herbivore defence genes in European aspen (Populus tremula L., Salicaceae). Molecular Ecology 21 (9):2197-2207. doi:10.1111/j.1365-294X.2012.05524.x

Berthelot A, Gavaland A (2007) Produire de la biomasse avec des taillis de peupliers. Informations-forêt 4 (760)

Billotte N, Jourjon MF, Marseillac N, Berger A, Flori A, Asmady H, Adon B, Singh R, Nouy B, Potier F, Cheah SC, Rohde W, Ritter E, Courtois B, Charrier A, Mangin B (2010) QTL detection by multi-parent linkage mapping in oil palm (Elaeis guineensis Jacq.). Theoretical and Applied Genetics 120 (8):1673-1687. doi:10.1007/s00122-010-1284-y

Bisoffi S, Gullberg U (1996) Poplar breeding and selection strategies. In: Stettler R, Heilman PE, Hinckley T, Bradshaw HD (eds) Biology of *Populus* and its implications for management and conservation. . NRC Research Press, pp 139-158.

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annual Review of Plant Biology 54:519-546. doi:10.1146/annurev.arplant.54.031902.134938

Bogdan S, Sporcic M, Seletkovic A, Ivankovic M (2009) Biomass Production of Common Alder (Alnus glutinosa/L./Gaertn.) in Pure Plantations and Mixed Plantations with Willow Clones (Salix sp.) in Croatia. Croatian Journal of Forest Engineering 30 (2):99-112

Bonduelle P Taillis à courtes rotations de peuplier (T.C.R.): Premiers résultats de développement en France. . In: al GGe (ed) Biomass for Energy and Industry, Fifth E.C. Conférence, Lisbon, Portugal, 1989. pp 1356–1360

Boyko A, Kovalchuk I (2011) Genome instability and epigenetic modification - heritable responses to environmental stress? Current Opinion in Plant Biology 14 (3):260-266. doi:10.1016/j.pbi.2011.03.003

Bradshaw HD, Ceulemans R, Davis J, Stettler R (2000) Emerging model systems in plant biology: Poplar (Populus) as a model forest tree. J Plant Growth Regul 19 (3):306-313. doi:10.1007/s003440000030

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

Bradshaw HD, Stettler R (1994) Molecular genetics of growth and development in Populus. II. Segregation distortion due to genetic load. Theor Appl Genet 89:551-558

Bradshaw HD, Stettler RF (1995) Molecular genetics of growth and development in Populus. 4. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. Genetics 139 (2):963-973

Breen AL, Glenn E, Yeager A, Olson MS (2009) Nucleotide diversity among natural populations of a North American poplar (Populus balsamifera, Salicaceae). New Phytologist 182 (3):763-773. doi:10.1111/j.1469-8137.2009.02779.x

Brereton NJB, Pitre FE, Hanley SJ, Ray MJ, Karp A, Murphy RJ (2010) QTL Mapping of Enzymatic Saccharification in Short Rotation Coppice Willow and Its Independence from Biomass Yield. Bioenergy Research 3 (3):251-261. doi:10.1007/s12155-010-9077-3

Brereton NJB, Pitre FE, Ray MJ, Karp A, Murphy RJ (2011) Investigation of tension wood formation and 2,6-dichlorbenzonitrile application in short rotation coppice willow composition and enzymatic saccharification. Biotechnology for Biofuels 4. doi:10.1186/1754-6834-4-13

Bresson A, Jorge V, Dowkiw A, Guerin V, Bourgait I, Tuskan GA, Schmutz J, Chalhoub B, Bastien C, Rampant PF (2011) Qualitative and quantitative resistances to leaf rust finely mapped within two nucleotide-binding site leucine-rich repeat (NBS-LRR)-rich genomic regions of chromosome 19 in poplar. New Phytologist 192 (1):151-163. doi:10.1111/j.1469-8137.2011.03786.x

Brignolas F, Paillassa E, Fichot R (2012a) FORINVEST, a forestry business angels? (Vers de nouveaux criteres de selection pour la populiculture dans un contexte de climat changeant.). Foret-Entreprise (204):59-64

Brignolas F, Paillassa E, Fichot R (2012b) New selection criteria for poplar in the context of changing climate (Vers de nouveaux criteres de selection pour la populiculture dans un contexte de climat changeant.). Bulletin Trimestriel - Centre de Populiculture du Hainaut (2):20-25

Brinkmann K, Blaschke L, Polle A (2002) Comparison of different methods for lignin determination as a basis for calibration of near-infrared reflectance spectroscopy and implications of lignoproteins. Journal of Chemical Ecology 28 (12):2483-2501. doi:10.1023/a:1021484002582

Britt C, Bullard M, Hickman G, Johnson P, King J, Nicholson F, Nixon P, Smith N (2002) Bioenergy crops and bioremediation - a review. A Contract Report by ADAS for the Department for Food, Environment and Rural Affairs

Broeckx LS, Verlinden MS, Ceulemans R (2012) Establishment and two-year growth of a bio-energy plantation with fast-growing Populus trees in Flanders (Belgium): Effects of genotype and former land use. Biomass Bioenerg 42:151-163. doi:10.1016/j.biombioe.2012.03.005

Brosche M, Vinocur B, Alatalo ER, Lamminmaki A, Teichmann T, Ottow EA, Djilianov D, Afif D, Bogeat-Triboulot MB, Altman A, Polle A, Dreyer E, Rudd S, Paulin L, Auvinen P, Kangasjarvi J (2005) Gene expression and metabolite profiling of Populus euphratica growing in the Negev desert. Genome Biology 6 (12):R101. doi:10.1186/gb-2005-6-12-r101

Brunner AM, Busov VB, Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends in Plant Science 9 (1):49-56. doi:10.1016/j.tplants.2003.11.006

Buhler DD, Netzer DA, Riemenschneider DE, Hartzler RG (1998) Weed management in short rotation poplar and herbaceous perennial crops grown for biofuel production. Biomass Bioenerg 14 (4):385-394. doi:10.1016/s0961-9534(97)10075-7

Bullard MJ, Mustill SJ, McMillan SD, Nixon PMI, Carver P, Britt CP (2002) Yield improvements through modification of planting density and harvest frequency in short rotation coppice Salix spp. - 1. Yield response in two morphologically diverse varieties. Biomass Bioenerg 22 (1):15-25. doi:10.1016/s0961-9534(01)00054-x

Busov V, Yordanov Y, Gou JQ, Meilan R, Ma C, Regan S, Strauss S (2011) Activation tagging is an effective gene tagging system in Populus. Tree Genetics & Genomes 7 (1):91-101. doi:10.1007/s11295-010-0317-7

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) Mixed models for S language environments, ASReml-R reference manual

Caffier V, Didelot F, Pumo B, Causeur D, Durel CE, Parisi L (2010) Aggressiveness of eight Venturia inaequalis isolates virulent or avirulent to the major resistance gene Rvi6 on a non-Rvi6 apple cultivar. Plant Pathol 59 (6):1072-1080. doi:10.1111/j.1365-3059.2010.02345.x

Campbell MM, Sederoff RR (1996) Variation in lignin content and composition - Mechanism of control and implications for the genetic improvement of plants. Plant Physiology 110 (1):3-13

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43 (10):956-963

Caslin B, Finnan J, McCracken A (2010) Willow best practice guidelines. 66pp. Teagasc, AFBI (ed).

Cervera MT, Gusmao J, Steenackers M, Storme V, Broeck Avd, Montagu Mv, Boerjan W (1996) The use of bulked segregant analysis to identify AFLP molecular markers closely linked to Melampsora larici-populina resistance in Populus. 205-210. Book: Somatic cell genetics and molecular genetics of trees.

Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Slycken Jv, 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

Ceulemans R, Scarascia Mugnozza GT, Wiard BM, Braatne JH, Hinckley TM, Stettler RF, Isebrands JG, Heilman PE (1992) Production physiology and morphology of Populus species and their hybrids grown under short rotation. I. Clonal comparisons of 4-year growth and phenology. Canadian Journal of Forest Research 22 (12):1937-1948

Chamaillard S (2011) Efficience d'utilisation de l'eau chez le peuplier noir (*Populus nigra* L.) : variabilité et plasticité en réponse aux variations de l'environnement. Université d'Orléans. Thèse de doctorat. 188 pp.

Chamaillard S, Fichot R, Vincent-Barbaroux C, Bastien C, Depierreux C, Dreyer E, Villar M, Brignolas F (2011) Variations in bulk leaf carbon isotope discrimination, growth and related leaf traits among three Populus nigra L. populations. Tree Physiology 31 (10):1076-1087. doi:10.1093/treephys/tpr089

Chang HM, Sarkanen KV (1973) Species variation in lignin. Effect of species on the rate of kraft delignification. Tappi 56 (3):132-134

Cheema J, Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. Briefings in Bioinformatics 10 (6):595-608. doi:10.1093/bib/bbp045

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nature Biotechnology 25 (7):759-761. doi:10.1038/nbt1316

Chen THH, Howe GT, Bradshaw HD, Jr. (2002) Molecular genetic analysis of dormancy-related traits in poplars. Weed Science 50 (2):232-240. doi:10.1614/0043-1745(2002)050[0232:mgaodr]2.0.co;2

Chiang VL, Funaoka M (1988) Comparaison of the responses of Guaiacyl and Guaiacyl-Syringyl nuclei to Kraft delignification. Abstracts of Papers of the American Chemical Society 195:205-CELL

Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, Hwang E-Y, Yi S-I, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176 (1):685-696. doi:10.1534/genetics.107.070821

Chu YG, Su XH, Huang QJ, Zhang XH (2009) Patterns of DNA sequence variation at candidate gene loci in black poplar (Populus nigra L.) as revealed by single nucleotide polymorphisms. Genetica 137 (2):141-150. doi:10.1007/s10709-009-9371-1

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138 (3):963-971

Cogan N, Smith K, Yamada T, Francki M, Vecchies A, Jones E, Spangenberg G, Forster J (2005) QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (Lolium perenne L.). Theoretical and Applied Genetics 110 (2):364-380. doi:10.1007/s00122-004-1848-9

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant Journal 37 (5):778-786. doi:10.1111/j.1365-313X.2003.01999.x

Dalianis, Christou M, Sooter C, Kyritsis S, Zafiris C, Samiotaki G (1994) Growth, yields and energy potential of densely planted Eucalyptusglobulus in a two year short rotation. Eucalyptus for Biomass Production (J S Pereira and H Pereira, Eds) Commission of the European Communities, Brussels/ Instituto Superior de Agronomia, Lisbon:193-201

Davison BH, Drescher SR, Tuskan GA, Davis MF, Nghiem NP (2006) Variation of S/G ratio and lignin content in a Populus family influences the release of xylose by dilute acid hydrolysis. Applied Biochemistry and Biotechnology 130 (1-3):427-435. doi:10.1385/abab:130:1:427

De Vienne D (1998) Les marqueurs moléculaires en génétique et biotechnologies végétales. 204pp. Quae (ed).

DeGomez T, Wagner MR (2001) Culture and use of black locust. HortTechnology 11 (2):279-288

Dejardin A, Leple JC, Lesage-Descauses MC, Costa G, Pilate G (2004) Expressed sequence tags from poplar wood tissues - A comparative analysis from multiple libraries. Plant Biology 6 (1):55-64. doi:10.1055/s-2003-44744

Demura T, Ye ZH (2010) Regulation of plant biomass production. Current Opinion in Plant Biology 13 (3):299-304. doi:10.1016/j.pbi.2010.03.002

Dereeper A, Nicolas S, Le Cunff L, Bacilieri R, Doligez A, Peros JP, Ruiz M, This P (2011) SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. Bmc Bioinformatics 12. doi:10.1186/1471-2105-12-134

Derozier S, Samson F, Tamby J-P, Guichard C, Brunaud V, Grevet P, Gagnot S, Label P, Leple J-C, Lecharny A, Aubourg S (2011) Exploration of plant genomes in the FLAGdb(++) environment. Plant Methods 7. doi:10.1186/1746-4811-7-8

Dharmawardhana P, Brunner AM, Strauss SH (2010) Genome-wide transcriptome analysis of the transition from primary to secondary stem development in Populus trichocarpa. BMC Genomics 11. doi:10.1186/1471-2164-11-150

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380 (6570):152-154. doi:10.1038/380152a0

Dickmann DI, Kuzovkina J (2008) Poplars and willows of the world, with emphasis on silviculturally important species. FAO (ed).

DiFazio S, Slavov G, Joshi C (2011) *Populus* : A Premier Pionner System for Plant Genomics. In: Joshi CD, S. & Kole, C. (ed) Genetics, Genomics and Breeding of poplar. Science Publishers, pp 1-28.

Dillen SY, El Kasmioui O, Marron N, Calfapietra C, Ceulemans R (2011) Poplar. In: Halford NG KA (ed) Energy crops. RSC Publishing, Cambridge, pp 275-300

Dillen SY, Rood SB, Ceulemans R (2010) Growth and Physiology. In: al. SJe (ed) Genetics and Genomics of Populus, Plant Genetics and Genomics: Crops and Models 8. Springer Science+Business Media, LLC 2010. doi:DOI 10.1007/978-1-4419-1541-2_3

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. doi:10.1007/s11295-008-0175-8

Dimitriou I, Rosenqvist H (2011) Sewage sludge and wastewater fertilisation of Short Rotation Coppice (SRC) for increased bioenergy production-Biological and economic potential. Biomass Bioenerg 35 (2):835-842. doi:10.1016/j.biombioe.2010.11.010

Dowkiw A, Bastien C (2004) Characterization of two major genetic factors controlling quantitative resistance to Melampsora larici-populina leaf rust in hybrid poplars: Strain specificity, field expression, combined effects, and relationship with a defeated qualitative resistance gene. Phytopathology 94 (12):1358-1367. doi:10.1094/phyto.2004.94.12.1358

Dowkiw A, Bastien C (2007) Presence of defeated qualitative resistance genes frequently has major impact on quantitative resistance to Melampsora larici-populina leaf rust in P.xinteramericana hybrid poplars. Tree Genetics & Genomes 3 (3):261-274. doi:10.1007/s11295-006-0062-0

Dowkiw A, Husson C, Frey P, Pinon J, Bastien C (2003) Partial resistance to Melampsora larici-populina leaf rust in hybrid poplars: genetic variability in inoculated excised leaf disk bioassay and relationship with complete resistance. Phytopathology 93 (4):421-427. doi:10.1094/phyto.2003.93.4.421

Dowkiw A, Voisin E, Bastien C (2010) Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. Plant Pathol 59 (3):523-534. doi:10.1111/j.1365-3059.2010.02277.x

Drost DR, Benedict CI, Berg A, Novaes E, Novaes CRDB, Yu Q, Dervinis C, Maia JM, Yap J, Miles B, Kirst M (2010) Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of Populus. Proceedings of the National Academy of Sciences of the United States of America 107 (18):8492-8497. doi:10.1073/pnas.0914709107

Drost DR, Benedict CI, Novaes E, Kirst M (2011) Populus genomic resources. In: Joshi CD, S. & Kole, C. (ed) Genetics, genomics and breeding of poplar. Science Publishers, pp 29-61.

Drost DR, Novaes E, Boaventura-Novaes C, Benedict CI, Brown RS, Yin T, Tuskan GA, Kirst M (2009) A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled Populus trichocarpa genome sequence. Plant Journal 58 (6):1054-1067. doi:10.1111/j.1365-313X.2009.03828.x

Du J, Groover A (2010) Transcriptional Regulation of Secondary Growth and Wood Formation. Journal of Integrative Plant Biology 52 (1):17-27. doi:10.1111/j.1744-7909.2010.00901.x

Dunlap JM, Stettler RF (1996) Genetic variation and productivity of Populus trichocarpa and its hybrids .9. Phenology and Melampsora rust incidence of native black cottonwood clones from four river valleys in Washington. For Ecol Manage 87 (1-3):233-256. doi:10.1016/s0378-1127(96)03774-7

Duplessis S, Hacquard S, Delaruelle C, Tisserant E, Frey P, Martin F, Kohler A (2011) Melampsora larici-populina transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. Molecular Plant-Microbe Interactions 24 (7):808-818. doi:10.1094/mpmi-01-11-0006

Duplessis S, Major I, Martin F, Seguin A (2009) Poplar and pathogen interactions: insights from Populus genome-wide analyses of resistance and defense gene families and gene expression profiling. Critical Reviews in Plant Sciences 28 (5):309-334. doi:10.1080/07352680903241063

Dupuis J, Siegmund D (1999) Statistical Methods for Mapping Quantitative Trait Loci From a Dense Set of Markers. Genetics 151 (1):373-386

Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V (2008) PlantGDB: a resource for comparative plant genomics. Nucleic Acids Research 36:D959-D965. doi:10.1093/nar/gkm1041

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464 (7291):1039-1042. doi:10.1038/nature08923

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6 (5). doi:10.1371/journal.pone.0019379

EU-Commission (2008) Communication : 20 20 by 2020, Europ's climate change opportunity.

Excoffier L, Lischer H (1995-2009) ARLEQUIN, An integrated Software Package for Population Genetics Data Analysis. In: Bioinformatics SIo (ed)

Fabbrini F, Gaudet M, Bastien C, Zaina G, Harfouche A, Beritognolo I, Marron N, Morgante M, Mugnozza GS, Sabatti M (2012) Phenotypic plasticity, QTL mapping and genomic characterization of bud set in black poplar. Bmc Plant Biology 12. doi:10.1186/1471-2229-12-47

FAO (1979) Poplars and willows in wood production and land use. vol XXXII, pp. 328.

FAO (2008) Synthèse des rapports d'activité nationaux, préparée pour la 23ème Session de la Commission internationale du peuplier, organisée conjointement par la FAO, et par la Beijing Forestry University, la State Forest Administration of China et la Chinese Academy of Forestry; Beijing, Chine. pp. 71.

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

FCBA (2009b) Mémento 2008-2009. pp. 35. www.fcba.fr.

FCBA (2012) Mémento 2012. pp. 41. www.fcba.fr.

Ferris R, Long L, Bunn SM, Robinson KM, Bradshaw HD, Rae AM, Taylor G (2002) Leaf stomatal and epidermal cell development: identification of putative quantitative trait loci in relation to elevated carbon dioxide concentration in poplar. Tree Physiology 22 (9):633-640

Fichot R, Barigah TS, Chamaillard S, Le Thiec D, Laurans F, Cochard H, Brignolas F (2010) Common trade-offs between xylem resistance to cavitation and other physiological traits do not hold among unrelated Populus deltoidesx Populus nigra hybrids. Plant Cell and Environment 33 (9):1553-1568. doi:10.1111/j.1365-3040.2010.02164.x

Fladung M, Buschbom J (2009) Identification of single nucleotide polymorphisms in different Populus species. Trees: Structure and Function 23 (6):1199-1212. doi:10.1007/s00468-009-0359-3

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54:357-374. doi:10.1146/annurev.arplant.54.031902.134907

Foston M, Hubbell CA, Samuel R, Jung S, Fan H, Ding SY, Zeng YN, Jawdy S, Davis M, Sykes R, Gjersing E, Tuskan GA, Kalluri U, Ragauskas AJ (2011) Chemical, ultrastructural and supramolecular analysis of tension wood in Populus tremula x alba as a model substrate for reduced recalcitrance. Energy Environ Sci 4 (12):4962-4971. doi:10.1039/c1ee02073k

Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet J-M, Cheynier V, This P (2009) Quantitative Genetic Bases of Anthocyanin Variation in Grape (Vitis vinifera L. ssp. sativa) Berry: A Quantitative Trait Locus to Quantitative Trait Nucleotide Integrated Study. Genetics 183 (3):1127-1139. doi:10.1534/genetics.109.103929

Fraga MF, Rodriguez R, Canal MJ (2002) Genomic DNA methylation-demethylation during aging and reinvigoration of Pinus radiata. Tree Physiology 22 (11):813-816

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu FL, Yang HM, Zeng CQ, Gao Y, Hu HR, Hu WT, Li CH, Lin W, Liu SQ, Pan H, Tang XL, Wang J, Wang W, Yu J, Zhang B, Zhang QR, Zhao HB, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao ZJ, Huang W, Chu X, He YG, Jin L, Liu YF, Shen YY, Sun WW, Wang HF, Wang Y, Wang Y, Xiong XY, Xu L, Waye MMY, Tsui SKW, Wong JTF, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai DM, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PKH, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PIW, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan WH, Li Y, Munro HM, Qin ZHS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Int HapMap C (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449 (7164):851-U853. doi:10.1038/nature06258

Freeman JS, Whittock SP, Potts BM, Vaillancourt RE (2009) QTL influencing growth and wood properties in Eucalyptus globulus. Tree Genetics & Genomes 5 (4):713-722. doi:10.1007/s11295-009-0222-0

Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD (2000) Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. Genetics 154 (2):837-845

Ganal MW, Altmann T, Roder MS (2009) SNP identification in crop plants. Current Opinion in Plant Biology 12 (2):211-217. doi:10.1016/j.pbi.2008.12.009

Gastine F, Berthelot A, Servant H, Roy B, Bouvet A (2003) Is the phytosanitary protection of the cultivar 'Beaupre' effective? Informations - Foret, Afocel (667):6 pp.

Gaudet M, Fara AG, Sabatti M, Kuztninsky E, Mugnozza GS (2007) Single-reaction for SNP genotyping on agarose gel by allele-specific PCR in black poplar (Populus nigra L.). Plant Molecular Biology Reporter 25 (1-2):1-9. doi:10.1007/s11105-007-0003-6

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. doi:10.1007/s11295-007-0085-1

Gendler K, Paulsen T, Napoli C (2008) ChromDB: The chromatin database. Nucleic Acids Research 36:D298-D302. doi:10.1093/nar/gkm768

Gergacz J (1978) Possibilities of breeding Aigeiros poplars for resistance. Agrartudomanyi Kozlemenyek 37 (2/3):271-274

Gerstch C, Moulin J, Dufloux C, B. V, Chopineau JL, Culan B, Désiré H, Nedelec H, Bodineau G (2012) Techniques sylvicoles et agricoles remédiantes, traitement et valorisation des effluents en circuit court par la production de biomasse. Projet TSAR.

Gibson G, Muse SV (2004) Précis de génomique. De Boeck (ed). 347 pp.

Gilchrist EJ, Haughn GW, Ying CC, Otto SP, Zhuang J, Cheung D, Hamberger B, Aboutorabi F, Kalynyak T, Johnson L, Bohlmann J, Ellis BE, Douglas CJ, Cronk QCB (2006) Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of Populus trichocarpa. Molecular Ecology 15 (5):1367-1378. doi:10.1111/j.1365-294X.2006.02885.x

Gion JM, Carouche A, Deweer S, Bedon F, Pichavant F, Charpentier JP, Bailleres H, Rozenberg P, Carocha V, Ognouabi N, Verhaegen D, Grima-Pettenati J, Vigneron P, Plomion C (2011) Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: Eucalyptus. BMC Genomics 12. doi:10.1186/1471-2164-12-301

Giordano E (1973) The genetic improvement of Poplar, with particular reference to the use of natural species and forms. Annali, Accademia Italiana di Scienze Forestali 22:159-181

Gonzalez-Garcia S, Martinez Gasol C, Teresa Moreira M, Gabarrell X, Rieradevall i Pons J, Feijoo G (2011) Environmental assessment of black locust (Robinia pseudoacacia L.)-based ethanol as potential transport fuel. International Journal of Life Cycle Assessment 16 (5):465-477. doi:10.1007/s11367-011-0272-z

Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in Pinus taeda L. Genetics 172 (3):1915-1926. doi:10.1534/genetics.105.047126

Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in Pinus taeda L. I. Wood property traits. Genetics 175 (1):399-409. doi:10.1534/genetics.106.061127

Goodstein DM, Shu SQ, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40 (D1):D1178-D1186. doi:10.1093/nar/gkr944

Goué N (2005) Etude de l'expression de gènes par type cellulaire dans le cambium de peuplier (*P. trichocarpa* x *P. deltoïdes* var Boelare). Thèse de doctorat. Université d'Orléans. 223 pp.

Gourcilleau D, Bogeat-Triboulot M-B, Le Thiec D, Lafon-Placette C, Delaunay A, Abu El-Soud W, Brignolas F, Maury S (2010) DNA methylation and histone acetylation: genotypic variations in hybrid poplars, impact of water deficit and relationships with productivity. Annals of Forest Science 67 (2). doi:10.1051/forest/2009101

Grattapaglia D, Plomion C, Kirst M, Sederoff RR (2009) Genomics of growth traits in forest trees. Current Opinion in Plant Biology 12 (2):148-156. doi:10.1016/j.pbi.2008.12.008

Grattapaglia D, Sederoff R (1994) Genetic-linkage maps of Eucalyptus-Grandis and Eucalyptus-Urophylla using a pseudo testcross - Mapping strategy and RAPD markers. Genetics 137 (4):1121-1137

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Research 36:D154-D158. doi:10.1093/nar/gkm952

Grunewald H, Bohm C, Quinkenstein A, Grundmann P, Eberts J, von Wuhlisch G (2009) Robinia pseudoacacia L.: A Lesser Known Tree Species for Biomass Production. Bioenergy Research 2 (3):123-133. doi:10.1007/s12155-009-9038-x

Guerra F, Wegrzyn JL, Sykes R, Davis J, Stanton BJ, Neale D (2013) Association genetics of chemical wood properties in black poplar (*Populus nigra* L.). New Phytologist 197 (1):162-176. doi:10.1111/nph.12003

Guet J (2011) Structuration géographique de la variabilité génétique de la phénologie de croissance et de l'efficience d'utilisation de l'eau chez le peuplier noir (*Populus nigra* L.). Master 2 Ecosystèmes Terrestres. Université d'Orléans. 28 pp.

Gustafson TJ, de Leon N (2010) Genetic Analysis of Maize (Zea mays L.) Endosperm Vitreousness and Related Hardness Traits in the Intermated B73 x Mo17 Recombinant Inbred Line Population. Crop Sci 50 (6):2318-2327. doi:10.2135/cropsci2009.12.0726

Hacquard S, Petre B, Frey P, Hecker A, Rouhier N, Duplessis S (2011) The poplar-poplar rust interaction: insights from genomics and transcriptomics. Journal of pathogens 2011:716041-716041. doi:10.4061/2011/716041

Hanley SJ, Mallott MD, Karp A (2007) Alignment of a Salix linkage map to the Populus genomic sequence reveals macrosynteny between willow and poplar genomes. Tree Genetics & Genomes 3 (1):35-48. doi:10.1007/s11295-006-0049-x

Hanley SJ, Pei MH, Powers SJ, Ruiz C, Mallott MD, Barker JHA, Karp A (2011) Genetic mapping of rust resistance loci in biomass willow. Tree Genetics and Genomes 7 (3):597-608. doi:10.1007/s11295-010-0359-x

Harrison EJ, Bush M, Plett JM, McPhee DP, Vitez R, O'Malley B, Sharma V, Bosnich W, Seguin A, MacKay J, Regan S (2007) Diverse developmental mutants revealed in an activation-tagged population of poplar. Canadian Journal of Botany-Revue Canadienne De Botanique 85 (11):1071-1081. doi:10.1139/b07-063

Heilman PE, Hinckley TM, Roberts DA, Ceulemans R (1996) Production physiology. In: Stettler R, Heilman PE, Hinckley T, Bradshaw HD (eds) Biology of *Populus* and its implications for management and conservation. NRC Research Press, pp 459-489.

Heilman PE, Stettler RF (1986) Nutritional concerns in selection of black cottonwood and hybrid clones for short rotation. Can J For Res-Rev Can Rech For 16 (4):860-863. doi:10.1139/x86-151

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A transcriptional roadmap to wood formation. Proceedings of the National Academy of Sciences of the United States of America 98 (25):14732-14737

Herve C, Ceulemans R (1996) Short-rotation coppiced vs non-coppiced poplar: A comparative study at two different field sites. Biomass Bioenerg 11 (2-3):139-150. doi:10.1016/0961-9534(96)00028-1

Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, Magrini VJ, Richt RJ, Sander SN, Stewart DA, Stromberg M, Tsung EF, Wylie T, Schedl T, Wilson RK, Mardis ER (2008) Whole-genome sequencing and variant discovery in C-elegans. Nat Methods 5 (2):183-188. doi:10.1038/nmeth.1179

Hinchee M, Rottmann W, Mullinax L, Zhang C, Chang S, Cunningham M, Pearson L, Nehra N (2009) Short-rotation woody crops for bioenergy and biofuels applications. In Vitro Cell Dev Biol-Plant 45 (6):619-629. doi:10.1007/s11627-009-9235-5

Hou S, Li L (2011) Rapid Characterization of Woody Biomass Digestibility and Chemical Composition Using Near-infrared SpectroscopyFree Access. Journal of Integrative Plant Biology 53 (2):166-175. doi:10.1111/j.1744-7909.2010.01003.x

Hu W, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai C, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nature Biotechnology 17 (8):808-812

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42 (11):961-967. doi:10.1038/ng.695

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44 (1):32-39. doi:10.1038/ng.1018

Hughes FMR, Rood SB (2003) Allocation of river flows for restoration of floodplain forest ecosystems: A review of approaches and their applicability in Europe. Environmental Management 31:12-33.

Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus, *Bremia-Lactucae*, using restriction fragment length polymorphisms. Genetics 120 (4):947-958

IFN (2011) Inventaire forestier national. www.ifn.fr. 16 pp. 28 (3e et 4e trimestre.)

Imbert E, Lefevre F (2003) Dispersal and gene flow of Populus nigra (Salicaceae) along a dynamic river system. Journal of Ecology 91 (3):447-456. doi:10.1046/j.1365-2745.2003.00772.x

Ingvarsson P, Garcia A, Hall D, Luquez V, Jansson S (2006) Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). Genetics 172:1845-1853. doi:10.1534/genetics.105.047522

Ingvarsson PK (2005a) Molecular population genetics of herbivore-induced protease inhibitor genes in European Aspen (Populus tremula L., Salicaceae). Molecular Biology and Evolution 22 (9):1802-1812. doi:10.1093/molbev/msi171

Ingvarsson PK (2005b) Nucleotide polymorphism and linkage disequilbrium within and among natural populations of European Aspen (Populus tremula L., Salicaceae). Genetics 169 (2):945-953. doi:10.1534/genetics.104.034959

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of Populus tremula. Genetics 180 (1):329-340. doi:10.1534/genetics.108.090431

Ingvarsson PK, Garcia MV, Luquez V, Hall D, Jansson S (2008) Nucleotide polymorphism and phenotypic associations within and around the phytochrome B2 locus in European aspen (Populus tremula, Salicaceae). Genetics 178 (4):2217-2226. doi:10.1534/genetics.107.082354

Isik F, Toplu F (2004) Variation in juvenile traits of natural black poplar (*Populus nigra* L.) clones in Turkey. New Forests 27:175-187

Ismail M, Soolanayakanahally RY, Ingvarsson PK, Guy RD, Jansson S, Silim SN, El-Kassaby YA (2012) Comparative Nucleotide Diversity Across North American and European Populus Species. Journal of Molecular Evolution 74 (5-6):257-272. doi:10.1007/s00239-012-9504-5

Jagtap SS, Dhiman SS, Jeya M, Kang YC, Choi J-H, Lee J-K (2012) Saccharification of poplar biomass by using lignocellulases from Pholiota adiposa. Bioresource technology 120:264-272. doi:http://dx.doi.org/10.1016/j.biortech.2012.06.002

Jansson S, Douglas CJ (2007) Populus: A model system for plant biology. In: Annual Review of Plant Biology, vol 58. Annual Review of Plant Biology. pp 435-458. doi:10.1146/annurev.arplant.58.032806.103956

Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology 57:19-53. doi:10.1146/annurev.arplant.57.032905.105218

Jones N, Ougham H, Thomas H, Pasakinskiene I (2009) Markers and mapping revisited: finding your gene. New Phytologist 183 (4):935-966. doi:10.1111/j.1469-8137.2009.02933.x

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. doi:10.1111/j.1469-8137.2005.01424.x

Jorgensen U, Dalgaard T, Kristensen ES (2005) Biomass energy in organic farming - the potential role of short rotation coppice. Biomass Bioenerg 28 (2):237-248. doi:10.1016/j.biombioe.2004.08.006

Kalluri U, Hurst GB, Lankford PK, Ranjan P, Pelletier DA (2009) Shotgun proteome profile of Populus developing xylem. Proteomics 9 (21):4871-4880. doi:10.1002/pmic.200800854

Karp A, Hanley SJ, Trybush SO, Macalpine W, Pei M, Shield I (2011) Genetic improvement of willow for bioenergy and biofuels. Journal of Integrative Plant Biology 53 (2):151-165. doi:10.1111/j.1744-7909.2010.01015.x

Karp A, Shield I (2008) Bioenergy from plants and the sustainable yield challenge. New Phytologist 179 (1):15-32. doi:10.1111/j.1469-8137.2008.02432.x

Kauter D, Lewandowski I, Claupein W (2003) Quantity and quality of harvestable biomass from Populus short rotation coppice for solid fuel use a review of the physiological basis and management influences. Biomass Bioenerg 24 (6):411-427. doi:10.1016/s0961-9534(02)00177-0

Kelleher CT, Wilkin J, Zhuang J, Cortés AJ, Quintero ALP, Gallagher TF, Bohlmann J, Douglas CJ, Ellis BE, Ritland K (2012) SNP discovery, gene diversity, and linkage disequilibrium in wild populations of Populus tremuloides. Tree Genetics & Genomes 8:821-829. doi:10.1007/s11295-012-0467-x

Keller SR, Levsen N, Ingvarsson PK, Olson MS, Tiffin P (2011) Local Selection Across a Latitudinal Gradient Shapes Nucleotide Diversity in Balsam Poplar, Populus balsamifera L. Genetics 188 (4):941-U318. doi:10.1534/genetics.111.128041

Keller SR, Olson MS, Silim S, Schroeder W, Tiffin P (2010) Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, Populus balsamifera. Molecular Ecology 19 (6):1212-1226. doi:10.1111/j.1365-294X.2010.04546.x

Kennard RW, Stone L (1969) Computer aided design of experiments. Technometrics 11:137-148

Kianian SF, Quiros CF (1992) Generation of a *Brassica-Oleracea* composite RFLP map - linkage arrangements among various populations end evolutionary implications. Theoretical and Applied Genetics 84 (5-6):544-554

Kim Y, Nielsen R (2004) Linkage Disequilibrium as a Signature of Selective Sweeps. Genetics 167 (3):1513-1524. doi:10.1534/genetics.103.025387

Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F (2008) Genome-wide identification of NBS resistance genes in Populus trichocarpa. Plant Molecular Biology 66 (6):619-636. doi:10.1007/s11103-008-9293-9

Köhn J (2009) Socio-economics in SRC – a review on concepts and the need for transdisciplinary research. Landbauforschung - vTI Agriculture and Forestry Research 59:223-232

Koskela J, De Vries SMG, Kajba D, Von Wuhlisch G (2004) Populus nigra network. Report of the seventh (25-27 October 2001, Osijek, Croatia) and eighth meetings (22-24 May 2003, Treppein, Germany) IPGRI, Rome, Italy

Krzan Z (1981) Poplar resistance to infection by the fungus Melampsora larici-populina in field conditions. Arboretum Kornickie 26:123-142

Labrecque M, Teodorescu TI (2005) Field performance and biomass production of 12 willow and poplar clones in short-rotation coppice in southern Quebec (Canada). Biomass Bioenerg 29 (1):1-9. doi:10.1016/j.biombioe.2004.12.004

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124 (3):743-756

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA (2009) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations (vol 1 pg 174, 1987). Genomics 93 (4):398-398. doi:10.1016/j.ygeno.2008.12.003

Lang T, Frigerio JM, Lalague H, Abadie P, Hubert F, Lepoittevin C, Decourcelle T, El Mujtar V, Léger V, Tibbits J, Paiva J, Kremer A, Franc A, Garnier-Gere P (2009) SeqQual: a user-friendly pipeline for integrating quality for identifying SNPs and producing sequence data files for population genetics analyses. . French Protection Agency Programs 2009. doi:IDDN.FR.001.380006.000.R.P.2009.000.10800

Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome. Am J Hum Genet 34 (6):842-845

Lapointe G, Luckevich MD, Cloutier M, Seguin A (2001) 14-3-3 gene family in hybrid poplar and its involvement in tree defence against pathogens. Journal of Experimental Botany 52 (359):1331-1338. doi:10.1093/jexbot/52.359.1331

Larsson S (1998) Genetic improvement of willow for short-rotation coppice. Biomass Bioenerg 15 (1):23-26. doi:10.1016/s0961-9534(98)80003-2

Lefèvre F, Achard P, Azais D, Smulders MJM, van der Schoot J, Bovenschen J, Ivens B, Storme V, Fluch S, Krystufek V, Castiglione S (2002) Distribution of Populus nigra genetic diversity within France and its consequences for ex-situ conservation strategy. In Genetic diversity in river populations of European black poplar. . Proceedings of an International Symposium held in Szeksárd, Hungary from 16-25 May, 2001 B C B van Dam, S Budapest, Hungary 85-92

Lefevre F, Legionnet A, Vries S, Turok J (1998) Strategies for the conservation of a pioneer tree species, Populus nigra L., in Europe. Genet Sel Evol 30:S181-S196. doi:10.1051/gse:19980711

Legionnet A, Muranty H, Lefevre F (1999) Genetic variation of the riparian pioneer tree species Populus nigra. II. Variation in susceptibility to the foliar rust Melampsora larici-populina. Heredity 82 (3):318-327. doi:10.1038/sj.hdy.6884880

Lemoine M, Pinon J (1978) Clonal differences in poplars in susceptibility to the rusts Melampsora larici-populina and M. allii-populina. Revue Forestiere Francaise (3):181-185

Leple JC, Dauwe R, Morreel K, Storme V, Lapierre C, Pollet B, Naumann A, Kang KY, Kim H, Ruel K, Lefebvre A, Joseleau JP, Grima-Pettenati J, De Rycke R, Andersson-Gunneras S, Erban A, Fehrle I, Petit-Conil M, Kopka J, Polle A, Messens E, Sundberg B, Mansfield SD, Ralph J, Pilate G, Boerjan W (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. doi:10.1105/tpc.107.054148

Lepoittevin C (2009) Génétique d'association chez le pin maritime (Pinus pinaster Ait.) pour la croissance et les composantes de la qualité du bois. Thèse de doctorat. Université de Bordeaux. 155 pp.

Lepoittevin C, Harvengt L, Plomion C, Garnier-Gere P (2012) Association mapping for growth, straightness and wood chemistry traits in the Pinus pinaster Aquitaine breeding population. Tree Genetics & Genomes 8 (1):113-126. doi:10.1007/s11295-011-0426-y

Li H, Liang Y, Xu Q, Cao D (2009) Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Analytica Chimica Acta 648 (1):77-84. doi:10.1016/j.aca.2009.06.046

Lin J, Gunter LE, Harding SA, Kopp RF, McCord RP, Tsai CJ, Tuskan GA, Smart LB (2007) Development of AFLP and RAPD markers linked to a locus associated with twisted growth in corkscrew willow (Salix matsudana 'Tortuosa'). Tree Physiology 27 (11):1575-1583

Liu Z, Furnier GR (1993) Inheritance and linkage of allozymes and restriction fragment length polymorphisms in trembling aspen. J Hered 84 (5):419-424

Lorne D, Chabrelie MF (2011) Les nouvelles technologies de production de biocarburants : état des lieux et enjeux des filières en développement. www.ifpen.fr. Panorama 2011 IFP énergies nouvelles

Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. Theoretical and Applied Genetics 105 (4):622-628. doi:10.1007/s00122-002-0970-9

Lu PL, Han XW, Qi J, Yang JG, Wijeratne AJ, Li T, Ma H (2012) Analysis of Arabidopsis genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg erecta and all four products of a single meiosis. Genome Res 22 (3):508-518. doi:10.1101/gr.127522.111

Lu SF, Sun YH, Chiang VL (2008) Stress-responsive microRNAs in Populus. Plant Journal 55 (1):131-151. doi:10.1111/j.1365-313X.2008.03497.x

Lu SF, Sun YH, Shi R, Clark C, Li LG, Chiang VL (2005) Novel and mechanical stress-responsive microRNAs in Populus trichocarpa that are absent from Arabidopsis. Plant Cell 17 (8):2186-2203. doi:10.1105/tpc.105.033456

Luger E (2000) Energy crop species in Europe. BLT Wieselburg, Austria

Lynch M, Walsh JB (1998) Genetics and Analysis of Quantitative Traits. Sinauer Assocs, Inc, Sunderland, MA:980 pp

Marron N, Storme V, Dillen SY, Bastien C, Ricciotti L, Salani F, Sabatti M, Rae AM, Ceulemans R, Boerjan W (2010) Genomic regions involved in productivity of two interspecific poplar families in Europe. 2. Biomass production and its relationships with tree architecture and phenology. Tree Genetics & Genomes 6 (4):533-554. doi:10.1007/s11295-010-0270-5

Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, Cattonaro F, Morgante M (2011a) Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. Plant Journal 67 (4):736-745. doi:10.1111/j.1365-313X.2011.04627.x

Marroni F, Pinosio S, Zaina G, Fogolari F, Felice N, Cattonaro F, Morgante M (2011b) Nucleotide diversity and linkage disequilibrium in Populus nigra cinnamyl alcohol dehydrogenase (CAD4) gene. Tree Genetics & Genomes 7 (5):1011-1023. doi:10.1007/s11295-011-0391-5

McCarthy RL, Zhong RQ, Fowler S, Lyskowski D, Piyasena H, Carleton K, Spicer C, Ye ZH (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. doi:10.1093/pcp/pcq064

Meier D, Fortmann I, Odermatt J, Faix O (2005) Discrimination of genetically modified poplar clones by analytical pyrolysis–gas chromatography and principal component analysis. Journal of Analytical and Applied Pyrolysis 74:129-137. doi:10.1016/j.jaap.2004.12.001

Melun F (2008) Espèces Ligneuses pour la production de biomasse : Eucalyptus. FCBA, Enerbio, Fondation Tuck. 4 pp.

Metzker ML (2010) Applications of next generation sequencing / Sequencing technologies - the next generation. Nature Reviews Genetics 11 (1):31-46. doi:10.1038/nrg2626

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 (4):1819-1829

Mevik BH, Wehrens R (2007) The pls Package: Principal Component and Partial Least Squares Regression in R. Journal of Statistical Software 18 (2)

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis - A rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United States of America 88 (21):9828-9832. doi:10.1073/pnas.88.21.9828

Miranda M, Ralph SG, Mellway R, White R, Heath MC, Bohlmann J, Constabel CP (2007) The transcriptional response of hybrid poplar (Populus trichocarpa x P-deltoides) to infection by Melampsora medusae leaf rust involves induction of flavonoid pathway genes leading to the accumulation of proanthocyanidins. Molecular Plant-Microbe Interactions 20 (7):816-831. doi:10.1094/mpmi-20-7-0816

Mitchell-Olds T, Bergelson J (2000) Biotic interactions Genomics and Coveolution. . Current Opinion in Plant Biology 3:273-277

Mitchell CP (1995) New cultural treatments and yield optimisation. Biomass Bioenerg 9 (1-5):11-34. doi:10.1016/0961-9534(95)00076-3

Mitchell CP, Stevens EA, Watters MP (1999) Short-rotation forestry - operations, productivity and costs based on experience gained in the UK. For Ecol Manage 121 (1-2):123-136. doi:10.1016/s0378-1127(98)00561-1

Monclus R, Dreyer E, Villar M, Delmotte FM, Delay D, Petit JM, Barbaroux C, Thiec D, Brechet C, Brignolas F (2006) Impact of drought on productivity and water use efficiency in 29 genotypes of Populus deltoides x Populus nigra. New Phytologist 169 (4):765-777. doi:10.1111/j.1469-8137.2005.01630.x

Monclus R, Leplé JC, Bastien C, Bert PF, Villar M, Marron N, Brignolas F, Jorge V (2012) QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp. Bmc Plant Biology in press

Monclus R, Villar M, Barbaroux C, Bastien C, Fichot R, Delmotte FM, Delay D, Petit JM, Brechet C, Dreyer E, Brignolas F (2009) Productivity, water-use efficiency and tolerance to moderate water deficit correlate in 33 poplar genotypes from a Populus deltoides x Populus trichocarpa F-1 progeny. Tree Physiology 29 (11):1329-1339. doi:10.1093/treephys/tpp075

Musso G, Costanzo M, Huangfu MQ, Smith AM, Paw J, Luis BJS, Boone C, Giaever G, Nislow C, Emili A, Zhang ZL (2008) The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. Genome Res 18 (7):1092-1099. doi:10.1101/gr.076174.108

Nanjo T, Futamura N, Nishiguchi M, Igasaki T, Shinozaki K, Shinohara K (2004) Characterization of full-length enriched expressed sequence tags of stress-treated poplar leaves. Plant and Cell Physiology 45 (12):1738-1748. doi:10.1093/pcp/pci009

Navarro D, Couturier M, da Silva GGD, Berrin J, Rouau X, Asther M, Bignon C (2010) Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. Microbial Cell Factories 9 (58)

Neale DB, Ingvarsson PK (2008) Population, quantitative and comparative genomics of adaptation in forest trees. Current Opinion in Plant Biology 11 (2):149-155. doi:10.1016/j.pbi.2007.12.004

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nature Reviews - Genetics 12 (2):111-122. doi:10.1038/nrg2931

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends in Plant Science 9 (7):325-330. doi:10.1016/j.plants.2004.05.006

Neale DB, Sewell MM, Brown GR (2002) Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. Annals of Forest Science 59 (5-6):595-605. doi:10.1051/forest:2002045

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America 76 (10):5269-5273. doi:10.1073/pnas.76.10.5269

Nepveu G, Keller R, Teissier du Cros E (1978) Sélection juvénile pour la qualité du bois chez certains peupliers noirs. Annales de Sciences Forestieres 35 (1):69-92

Newcombe G (1998) A review of exapted resistance to diseases of Populus. European Journal of Forest Pathology 28:209-216

Newcombe G, Bradshaw HD (1996) Quantitative trait loci conferring resistance in hybrid poplar to Septoria populicola, the cause of leaf spot. Can J For Res-Rev Can Rech For 26 (11):1943-1950. doi:10.1139/x26-219

Newcombe G, Bradshaw HD, Chastagner GA, Stettler RF (1996) A major gene for resistance to Melampsora medusae f sp deltoidae in a hybrid poplar pedigree. Phytopathology 86 (1):87-94. doi:10.1094/Phyto-86-87

Novaes E, Kirst M, Chiang V, Winter-Sederoff H, Sederoff R (2010) Lignin and Biomass: A Negative Correlation for Wood Formation and Lignin Content in Trees. Plant Physiology 154 (2):555-561. doi:10.1104/pp.110.161281

Novaes E, Osorio L, Drost DR, Miles BL, Boaventura-Novaes CRD, Benedict C, Dervinis C, Yu Q, Sykes R, Davis M, Martin TA, Peter GF, Kirst M (2009) Quantitative genetic analysis of biomass and wood chemistry of Populus under different nitrogen levels. New Phytologist 182 (4):878-890. doi:10.1111/j.1469-8137.2009.02785.x

Noveltree (2012) Proceeding of noveltree final conference "Tree breeding, genomics and evolutionary Biology : new synergies to tackle the impact of climate change in the 21th century". 17-18 October, Helsinki.

Ohlrogge J, Allen D, Berguson B, DellaPenna D, Shachar-Hill Y, Stymne S (2009) Driving on Biomass. Science 324 (5930):1019-1020. doi:10.1126/science.1171740

Olson MS, Robertson AL, Takebayashi N, Silim S, Schroeder WR, Tiffin P (2010) Nucleotide diversity and linkage disequilibrium in balsam poplar (Populus balsamifera). New Phytologist 186 (2):526-536. doi:10.1111/j.1469-8137.2009.03174.x

Pakull B, Groppe K, Meyer M, Markussen T, Fladung M (2009) Genetic linkage mapping in aspen (Populus tremula L. and Populus tremuloides Michx.). Tree Genetics and Genomes 5 (3):505-515. doi:10.1007/s11295-009-0204-2

Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E, Muleo R, Mugnozza GS, Sabatti M (2010) Genetic linkage maps of Populus alba L. and comparative mapping analysis of sex determination across Populus species. Tree Genetics & Genomes 6 (6):863-875. doi:10.1007/s11295-010-0297-7

Park S, Yu H-J, Mun J-H, Lee S-C (2010) Genome-wide discovery of DNA polymorphism in Brassica rapa. Molecular Genetics and Genomics 283 (2):135-145. doi:10.1007/s00438-009-0504-0

Pei M, Hunter T, Publishing BIPC (2005) Disease scoring by taking inoculum densities into consideration in leaf disc inoculations with poplar and willow rusts. Rust diseases of willow and poplar. doi:10.1079/9780851999999.0131

Persillet V (2012) Les biocarburants de première génération : un bilan mondial mitigé (version révisée Juillet 2012). INRA Sciences Sociales (1/2012)

Perttu KL (1998) Environmental justification for short-rotation forestry in Sweden. Biomass Bioenerg 15 (1):1-6. doi:10.1016/s0961-9534(98)00014-2

Petersen JE (2007) Short Rotation Forestry, Short Rotation Coppice andenergy grasses in the European Union: Agro-environmental aspects, present use and perspectives. EEA Specific Contract No 2 3604/B2006/EEA52793 - Deliverable Task 2b: Background report on current SRF/SRC cropping patterns in Europe

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7 (4):275-291. doi:10.1023/a:1011605013259

Philippe RN, Ralph SG, Kulheim C, Jancsik SI, Bohlmann J (2009) Poplar defense against insects: genome analysis, full-length cDNA cloning, and transcriptome and protein analysis of the poplar Kunitz-type protease inhibitor family. New Phytologist 184 (4):865-884. doi:10.1111/j.1469-8137.2009.03028.x

Philippe RN, Ralph SG, Mansfield SD, Bohlmann J (2010) Transcriptome profiles of hybrid poplar (Populus trichocarpa x deltoides) reveal rapid changes in undamaged, systemic sink leaves after simulated feeding by forest tent caterpillar (Malacosoma disstria). New Phytologist 188 (3):787-802. doi:10.1111/j.1469-8137.2010.03392.x

Pichot C, Teissier du Cros E (1988) Estimation of genetic-parameters in the European black poplar (*Populus-nigra L.*) - Consequence on the breeding strategy. Ann Sci For 45 (3):223-237. doi:10.1051/forest:19880304

Pinon J (1992) Variability in the genus *Populus* in sensitivity to *Melampsora* rusts. Silvae Genetica 41 (1):25-34

Pinon J, Du Cros ET (1976) Susceptibility to rusts in miscellaneous poplar species Melampsora larici-populina Kleb. and M. allii-populina Kleb. Annales de Sciences Forestieres 33 (2):49-59

Pinon J, Frey P (2005) Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance. Rust Diseases of Willow and Poplar M H Pei and A R McCracken, eds CAB International, Wallingford, UK:139-154

Pinon J, Schvester D (1985) Problèmes phytosanitaires dans les taillis à courte rotation. Revue forestière française XXVII:17-25

Pitre FE, Brereton NJB, Audoire S, Richter GM, Shield I, Karp A (2010) Estimating root biomass in Salix viminalis x Salix schwerinii cultivar oOlofo using the electrical capacitance method. Plant Biosystems 144 (2):479-483. doi:10.1080/11263501003732092

Plomion C, Leprovost G, Stokes A (2001) Wood formation in trees. Plant Physiology 127 (4):1513-1523. doi:10.1104/pp.010816

Prat D, Faivre-Rampant P, Prado E (2006) Analyse du génome et gestion des ressources génétiques forestières. INRA Quae (ed). 484 pp.

Puri VP (1984) Effect of cristallinity and degree of polymerization of cellulose on enzymatic saccharification. Biotechnol Bioeng 26 (10):1219-1222. doi:10.1002/bit.260261010

Rae AM, Ferris R, Tallis MJ, Taylor G (2006) Elucidating genomic regions determining enhanced leaf growth and delayed senescence in elevated CO2. Plant Cell and Environment 29 (9):1730-1741. doi:10.1111/j.1365-3040.2006.01545.x

Rae AM, Pinel MPC, Bastien C, Sabatti M, Street NR, Tucker J, Dixon C, Marron N, Dillen SY, Taylor G (2008) QTL for yield in bioenergy Populus: identifying G * E interactions from growth at three contrasting sites. Tree Genetics and Genomes 4 (1):97-112. doi:10.1007/s11295-007-0091-3

Rae AM, Street NR, Robinson KM, Harris N, Taylor G (2009) Five QTL hotspots for yield in short rotation coppice bioenergy poplar: the Poplar Biomass Loci. Bmc Plant Biology 9 (23). doi:10.1186/1471-2229-9-23

Rae AM, Tricker PJ, Bunn SM, Taylor G (2007) Adaptation of tree growth to elevated CO2: quantitative trait loci for biomass in Populus. New Phytologist 175 (1):59-69. doi:10.1111/j.1469-8137.2007.02091.x

Rafalski JA (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162 (3):329-333. doi:10.1016/s0168-9452(01)00587-8

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311 (5760):484-489. doi:10.1126/science.1114736

Ralph J, Akiyama T, Kim H, Lu FC, Schatz PF, Marita JM, Ralph SA, Reddy MSS, Chen F, Dixon RA (2006) Effects of coumarate 3-hydroxylase down-regulation on lignin structure. Journal of Biological Chemistry 281 (13):8843-8853. doi:10.1074/jbc.M511598200

Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, Boerjan W (2004) Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. . Phytochemistry Reviews 3:29–60

Ralph SG, Chun HJE, Cooper D, Kirkpatrick R, Kolosova N, Gunter L, Tuskan GA, Douglas CJ, Holt RA, Jones SJM, Marra MA, Bohlmann J (2008) Analysis of 4,664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding. BMC Genomics 9. doi:10.1186/1471-2164-9-57

Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (Sorghum bicolor L. Moench). Theoretical and Applied Genetics 97 (4):605-616. doi:10.1007/s001220050936

Ray MJ, Brereton NJB, Shield I, Karp A, Murphy R (2012) Variation in Cell Wall Composition and Accessibility in Relation to Biofuel Potential of Short Rotation Coppice Willows. Bioenergy Research 5:685-698. doi:10.1007/s12155-011-9177-8

Redei K (1998) Black locust (Robinia pseudoacacia L.) growing in Hungary. . Publications of the Hungarian Forest Research Institute, Budapest, Ungarn

Redei K, Csiha I, Keseru Z, Raso J (2009) Initial spacing effects on the stand structure factors in young black locust (Robinia pseudoacacia L.) stands. Lesnicky Casopis 55 (4):395-400

Resende MDV, Resende MFR, Jr., Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, Jr., Kilian A, Grattapaglia D (2012a) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist 194 (1):116-128. doi:10.1111/j.1469-8137.2011.04038.x

Resende MFR, Jr., Munoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012b) Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (Pinus taeda L.). Genetics 190 (4):1503-+. doi:10.1534/genetics.111.137026

Riano-Pachon DM, Nagel A, Neigenfind J, Wagner R, Basekow R, Weber E, Mueller-Roeber B, Diehl S, Kersten B (2009) GabiPD: the GABI primary database-a plant integrative 'omics' database. Nucleic Acids Research 37:D954-D959. doi:10.1093/nar/gkn611

Riemenschneideir DE, Stanton BJ, Vallée G, Périnet P (2001) Poplar breeding strategies. In: Dickmann DI, Isebrands JG, Eckenwalder JE, Richardson J (eds) Poplar culture in North America. NRC Research Press, pp 43-76.

Riemenschneideir DE, Stelzer HE, Foster GS (1996) Quantitative genetics of poplars and poplar hybrids. In: Stettler R, Heilman PE, Hinckley T, Bradshaw HD (eds) Biology of *Populus* and its implications for management and conservation. NRC Research Press, pp 159-181.

Rime D, Robert C, Goyeau H, Lannou C (2005) Effect of host genotype on leaf rust (Puccinia triticina) lesion development and urediniospore production in wheat seedlings. Plant Pathol 54 (3):287-298. doi:10.1111/j.1365-3059.2005.01174.x

Rinaldi C, Kohler A, Frey P, Duchaussoy F, Ningre N, Couloux A, Wincker P, Thiec Dl, Fluch S, Martin F, Duplessis S (2007) Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust Melampsora larici-populina. Plant Physiology 144 (1):347-366. doi:10.1104/pp.106.094987

Robinson AR, Mansfield SD (2009) Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. Plant Journal 58 (4):706-714. doi:10.1111/j.1365-313X.2009.03808.x

Rohde A, Storme V, Jorge V, Gaudet M, Vitacolonna N, Fabbrini F, Ruttink T, Zaina G, Marron N, Dillen S, Steenackers M, Sabatti M, Morgante M, Boerjan W, Bastien C (2011) Bud set in poplar - genetic dissection of a complex trait in natural and hybrid populations. New Phytologist 189 (1):106-121. doi:10.1111/j.1469-8137.2010.03469.x

Ronnberg-Wastljung AC, Tsarouhas V, Semerikov V, Lagercrantz U (2003) A genetic linkage map of a tetraploid Salix viminalis * S. dasyclados hybrid based on AFLP markers. Forest Genetics 10 (3):185-194

Rubin EM (2008) Genomics of cellulosic biofuels. Nature 454 (7206):841-845. doi:10.1038/nature07190

Ruffinoni C, Gazelle F (1997) Riparian forests and alluvial forests: restoration and management (Ripisylves et forets alluviales: restauration et gestion des ripisylves). Foret-Entreprise (116):43-49

Samils B, Ronnberg-Wastljung AC, Stenlid J (2011) QTL mapping of resistance to leaf rust in Salix. Tree Genetics & Genomes 7 (6):1219-1235. doi:10.1007/s11295-011-0408-0

Sannigrahi P, Ragauskas AJ, Tuskan GA (2010) Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuels, Bioproducts & Biorefining 4 (2):209-226. doi:10.1002/bbb.206

Savitzky A, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36:1627

Savolainen O, Pyhajarvi T (2007) Genomic diversity in forest trees. Current Opinion in Plant Biology 10 (2):162-167. doi:10.1016/j.pbi.2007.01.011

Scarascia-Mugnozza GE, Hinckley TM, Stettler RF, Heilman PE, Isebrands JG (1999) Production physiology and morphology of Populus species and their hybrids grown under short rotation. III. Seasonal carbon allocation patterns from branches. Can J For Res-Rev Can Rech For 29 (9):1419-1432. doi:10.1139/cjfr-29-9-1419

Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B (2003) Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana. Genome Res 13 (6):1250-1257. doi:10.1101/gr.728603

Schoeneman RS (1986) Cultural practices and leaf rust resistance in integrated pest management of Populus plantations. Dissertation Abstracts International, B (Sciences and Engineering) 47 (5):1813-B

Schoot Jvd, Pospiskova M, Vosman B, Smulders MJM (2000) Development and characterization of microsatellite markers in black poplar (Populus nigra L.). Theoretical and Applied Genetics 101 (1/2):317-322. doi:10.1007/s001220051485

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18 (2):233-234. doi:10.1038/72708

Searchinger T, Heimlich R, Houghton RA, Dong FX, Elobeid A, Fabiosa J, Tokgoz S, Hayes D, Yu TH (2008) Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319 (5867):1238-1240. doi:10.1126/science.1151861

Semerikov V, Lagercrantz U, Tsarouhas V, Ronnberg-Wastljung A, Alstrom-Rapaport C, Lascoux M (2003) Genetic mapping of sex-linked markers in Salix viminalis L. Heredity 91 (3):293-299. doi:10.1038/sj.hdy.6800327

Sennerbyforsse L, Zsuffa L (1995) Bud structure and resprouting in coppiced stools of Salix-Viminalis L, S-Eriocephala Michx, and S-Amygdaloides Anders. Trees-Structure and Function 9 (4):224-234

Sjodin A, Street NR, Sandberg G, Gustafsson P, Jansson S (2009) The Populus Genome Integrative Explorer (PopGenIE): a new resource for exploring the Populus genome. New Phytologist 182 (4):1013-1025. doi:10.1111/j.1469-8137.2009.02807.x

Skipars V, Baumanis I, Rungis D (2010) Research on genetic aspects of Scots pine resistance to root rot. Annual 16th International Scientific Conference Proceedings, "Research for rural development 2010", Jelgava, Latvia, 19-21 May 2010 Volume 1:222-226

Slatkin M (2008) Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9 (6):477-485. doi:10.1038/nrg2361

Slavov GT, DiFazio SP, Martin J, Schackwitz W, Muchero W, Rodgers-Melnick E, Lipphardt MF, Pennacchio CP, Hellsten U, Pennacchio LA, Gunter LE, Ranjan P, Vining K, Pomraning KR, Wilhelm LJ, Pellegrini M, Mockler TC, Freitag M, Geraldes A, El-Kassaby YA, Mansfield SD, Cronk QCB, Douglas CJ, Strauss SH, Rokhsar D, Tuskan GA (2012) Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. New Phytologist 196 (3):713-725. doi:10.1111/j.1469-8137.2012.04258.x

Slycken Jv, Michiels B, Steenackers M (2002) New poplar clones developed at IBW: 'Muur', 'Vesten', 'Oudenberg' and 'Grimminge'. Bulletin Trimestriel - Centre de Populiculture du Hainaut (4):67-83

Smulders MJM, Beringen R, Volosyanchuk R, Broeck AV, van der Schoot J, Arens P, Vosman B (2008a) Natural hybridisation between Populus nigra L. and P. x canadensis Moench. Hybrid offspring competes for niches along the Rhine river in the Netherlands. Tree Genetics & Genomes 4 (4):663-675. doi:10.1007/s11295-008-0141-5

Smulders MJM, Cottrell JE, Lefevre F, van der Schoot J, Arens P, Vosman B, Tabbener HE, Grassi F, Fossati T, Castiglione S, Krystufek V, Fluch S, Burg K, Vornam B, Pohl A, Gebhardt K, Alba N, Agundez D, Maestro C, Notivol E, Volosyanchuk R, Pospiskova M, Bordacs S, Bovenschen J, van Dam BC, Koelewijn HP, Halfmaerten D, Ivens B, van Slycken J, Broeck AV, Storme V, Boerjan W (2008b) Structure of the genetic diversity in black poplar (Populus nigra L.) populations across European river systems: Consequences for conservation and restoration. For Ecol Manage 255 (5-6):1388-1399. doi:10.1016/j.foreco.2007.10.063

Spinelli R, Nati C, Magagnotti N (2008) Harvesting Short-Rotation Poplar Plantations for Biomass Production. Croatian Journal of Forest Engineering 29 (2):129-139

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant Journal 3 (5):739-744. doi:10.1046/j.1365-313X.1993.03050739.x

Stanturf JA, Van Oosten C, Netzer DA, Coleman MD, Portwood CJ (2001) Ecology and silviculture of poplar plantations. In: Dickmann DI, Isebrands JG, Eckenwalder JE, Richardson J (eds) Poplar culture in North America. NRC Research Press, pp 153-206

Steemers FJ, Chang WH, Lee G, Barker DL, Shen R, Gunderson KL (2006) Whole-genome genotyping with the single-base extension assay. Nat Methods 3 (1):31-33. doi:10.1038/nmeth842

Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH, Sundberg B, Gustafsson P, Uhlen M, Bhalerao RP, Nilsson O, Sandberg G,

Karlsson J, Lundeberg J, Jansson S (2004) A Populus EST resource for plant functional genomics. Proceedings of the National Academy of Sciences of the United States of America 101 (38):13951-13956. doi:10.1073/pnas.0401641101

Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, Van Montagu M, Sandberg G, Olsson O, Teeri TT, Boerjan W, Gustafsson P, Uhlen M, Sundberg B, Lundeberg J (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. doi:10.1073/pnas.95.22.13330

Stettler R, Zsuffa L, Wu R (1996) The role of hybridisation in the genetic manipulation of *Populus* In: Stettler R, Heilman PE, Hinckley T, Bradshaw HD (eds) Biology of *Populus* and its implications for management and conservation. pp 87-112.

Stettler RF, Fenn RC, Heilman PE, Stanton BJ (1988) *Populus trichocarpa* x *Populus deltoides* hybrids for short rotation culture: Variation patterns and 4-year field performance. Canadian Journal of Forest Research 18:745-753

Stirling B, Newcombe G, Vrebalov J, Bosdet I, Bradshaw HD (2001) Suppressed recombination around the MXC3 locus, a major gene for resistance to poplar leaf rust. Theoretical and Applied Genetics 103 (8):1129-1137. doi:10.1007/s001220100721

Stolarski M, Szczukowski S, Tworkowski J, Klasa A (2008) Productivity of seven clones of willow coppice in annual and quadrennial cutting cycles. Biomass Bioenerg 32 (12):1227-1234. doi:10.1016/j.biombioe.2008.02.023

Storme V, Vanden Broeck A, Ivens B, Halfmaerten D, Van Slycken J, Castiglione S, Grassi F, Fossati T, Cottrell JE, Tabbener HE, Lefevre F, Saintagne C, Fluch S, Krystufek V, Burg K, Bordacs S, Borovics A, Gebhardt K, Vornam B, Pohl A, Alba N, Agundez D, Maestro C, Notivol E, Bovenschen J, van Dam BC, van der Schoot J, Vosman B, Boerjan W, Smulders MJM (2004) Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development. Theoretical and Applied Genetics 108 (6):969-981. doi:10.1007/s00122-003-1523-6

Street NR, Skogstrom O, Sjodin A, Tucker J, Rodriguez-Acosta M, Nilsson P, Jansson S, Taylor G (2006) The genetics and genomics of the drought response in Populus. Plant Journal 48 (3):321-341. doi:10.1111/j.1365-313X.2006.02864.x

Street NR, Tallis MJ, Tucker J, Brosche M, Kangasjarvi J, Broadmeadow M, Taylor G (2011) The physiological, transcriptional and genetic responses of an ozone- sensitive and an ozone tolerant poplar and selected extremes of their F-2 progeny. Environmental Pollution 159 (1):45-54. doi:10.1016/j.envpol.2010.09.027

Strong T (1989) Rotation Length and Repeated Harvesting Forest Service Influence Populus Coppice Production. United States Department of Agriculture. Research Note NC-350.

Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE (2011) Lignin content in natural Populus variants affects sugar release. Proceedings of the National Academy of Sciences of the United States of America 108 (15):6300-6305. doi:10.1073/pnas.1009252108

Su X, Zhang Q, Zheng X, Zhang X, Harrish S (1998) Construction of Populus deltoides Marsh * P. cathayana Rehd. molecular linkage map. Scientia Silvae Sinicae 34 (6):29-37

Sykes R, Yung M, Novaes E, Kirst M, Peter G, Davis M (2009) High-throughput screening of plant cell-wall composition using pyrolysis molecular beam mass spectroscopy. Methods in molecular biology (Clifton, NJ) 581:169-183

Sylvabiom (2008 - 2012) Projet ANR :Nouveaux concepts de cultures ligneuses durables pour la production de biomasse à des fins énergétiques.

Takahashi M, Matsuda F, Margetic N, Lathrop M (2003) Automated Identification of Single Nucleotide Polymorphisms from Sequencing Data. J Bioinform Comput Biol 1 (2):253-265

Taleei A, Siahsar BA, Peighambari SA (2009) QTL Mapping for Forage Quality-Related Traits in Barley. In: Slezak D, Arslan T, Fang WC, Song X, Kim TH (eds) Bio-Science and Bio-Technology, vol 57. Communications in Computer and Information Science. pp 53-62

TAPPI (1991) TAPPI standard UM250. Acid-soluble lignin in wood and pulp. TAPPI J, Atlanta, GA

Taylor G (2002) *Populus*: Arabidopsis for forestry. Do we need a model tree? Annals of Botany 90 (6):681-689. doi:10.1093/aob/mcf255

Teissier du Cros E (1977) Aperçu de la transmission héréditaire de quelques caractères juvéniles chez *populus nigra* L. Annales de Sciences Forestieres 34 (4):311-322

Thamarus K, Groom K, Bradley A, Raymond CA, Schimleck LR, Williams ER, Moran GF (2004) Identification of quantitative trait loci for wood and fibre properties in two full-sib pedigrees of Eucalyptus globulus. Theoretical and Applied Genetics 109 (4):856-864. doi:10.1007/s00122-004-1699-4

Thammasouk K, Tandjo D, Penner MH (1997) Influence of extractives on the analysis of herbaceous biomass. Journal of Agricultural and Food Chemistry 45 (2):437-443. doi:10.1021/jf960401r

Tharakan PJ, Volk TA, Abrahamson LP, White EH (2003) Energy feedstock characteristics of willow and hybrid poplar clones at harvest age. Biomass Bioenerg 25 (6):571-580. doi:10.1016/s0961-9534(03)00054-0

Theander O, Aman P, Westerlund E, Andersson R, Petersson D (1995) Total dietary fiber determined as neutral sugar residues, uronic acid residues, and Klason Lignin (The Uppsala method): Collaborative study. Journal of Aoac International 78 (4):1030-1044

Thumma BR, Baltunis BS, Bell JC, Emebiri LC, Moran GF, Southerton SG (2010) Quantitative trait locus (QTL) analysis of growth and vegetative propagation traits in Eucalyptus nitens full-sib families. Tree Genetics & Genomes 6 (6):877-889. doi:10.1007/s11295-010-0298-6

Thumma BR, Matheson BA, Zhang DQ, Meeske C, Meder R, Downes GM, Southerton SG (2009) Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene Affecting Cellulose Content. Genetics 183 (3):1153-1164. doi:10.1534/genetics.109.106591

Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in Eucalyptus spp. Genetics 171 (3):1257-1265. doi:10.1534/genetics.105.042028

Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM, Wu PP, Wang Y, Spoonde AY, Koehler RT, Peyret N, Chen C, Broomer AJ, Ridzon DA, Zhou H, Hoo BS, Hayashibara KC, Leong LN, Ma CN, Rosenblum BB, Day JP, Ziegle JS, De La Vega FM, Rhodes MD, Hennessy KM, Wenz HM (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. Journal of biomolecular techniques : JBT 16 (4):398-406

Torjek O, Witucka-Wall H, Meyer RC, von Korff M, Kusterer B, Rautengarten C, Altmann T (2006) Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theoretical and Applied Genetics 113 (8):1551-1561. doi:10.1007/s00122-006-0402-3

Tsai C, Popko JL, Mielke MR, Hu W, Podila GK, Chiang VL (1998) Suppression of O-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown wood phenotypes. Plant Physiology 117 (1):101-112. doi:10.1104/pp.117.1.101

Tsai CJ, Ranjan P, diFazio S, Tuskan G, Johnson VE (2011) Poplar Genome Microarrays. In: Joshi CD, S. & Kole, C. (ed) Genetics, Genomics and Breeding of Poplar. Science Publishers, pp 112-127.

Tschaplinski TJ, Tuskan GA, Sewell MM, Gebre GM, Donald ETI, Pendleyi C (2006) Phenotypic variation and quantitative trait locus identification for osmotic potential in an interspecific hybrid inbred F-2 poplar pedigree grown in contrasting environments. Tree Physiology 26 (5):595-604

Tuskan GA, 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 RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, 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 CJ, 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 (5793):1596-1604. doi:10.1126/science.1128691

Tuskan GA, DiFazio SP, Faivre-Rampant P, Gaudet M, Harfouche A, Jorge V, Labbé JL, Ranjan P, Sabatti M, Slavov G, Street NR, Tschaplinski TJ, Yin T (2012) The obscure events contributing to the evolution of an incipient sex chromosome in Populus : a retrospective working hypothesis. Tree Genetics & Genomes. doi:10.1007/s11295-012-0495-6

Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM, DiFazio SP (2004) Characterization of microsatellites revealed by genomic sequencing of Populus trichocarpa. Canadian Journal of Forest Research 34 (1):85-93. doi:10.1139/x03-283

Uri V, Lohmus K, Kiviste A, Aosaar J (2009) The dynamics of biomass production in relation to foliar and root traits in a grey alder (Alnus incana (L.) Moench) plantation on abandoned agricultural land. Forestry 82 (1):61-74. doi:10.1093/forestry/cpn040

Uri V, Lohmus K, Mander U, Ostonen I, Aosaar J, Maddison M, Helmisaari HS, Augustin J (2011) Long-term effects on the nitrogen budget of a short-rotation grey alder (Alnus incana (L.) Moench) forest on abandoned agricultural land. Ecological Engineering 37 (6):920-930. doi:10.1016/j.ecoleng.2011.01.016

Van Dam BC, Bordacs S (2002) Genetic diversity in river populations of European black poplar. Proceedings of an International Symposium held in Szeksárd, Hungary from 16-25 May, 2001 Csiszár Nyomda, Budapest, Hungary

Van Ooijen JW (2006) JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherrlands

Van Ooijen JW (2009) MapQTL 6, software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen, The Netherlands

Van Soest PJ, Robertson JD, Lewis BA (1991) Methods for dietary fiber, neutral detergent fiber and non-starch polysaccharide in relation to animal nutrition. J Dairy Science 74:3583-3597

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin Biosynthesis and Structure. Plant Physiology 153 (3):895-905. doi:10.1104/pp.110.155119

Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Current Opinion in Plant Biology 11 (3):278-285. doi:10.1016/j.pbi.2008.03.005

Varshney RK, Glaszmann J-C, Leung H, Ribaut J-M (2010) More genomic resources for less-studied crops. Trends Biotechnol 28 (9):452-460. doi:10.1016/j.tibtech.2010.06.007

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genetics, Selection, Evolution 34 (3):275-305. doi:10.1051/gse:2002009

Villar M, Lefevre F, Bradshaw HD, duCros ET (1996) Molecular genetics of rust resistance in poplars (Melampsora larici-populina Kleb Populus sp) by bulked segregant analysis in a 2x2 factorial mating design. Genetics 143 (1):531-536

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93 (1):77-78. doi:10.1093/jhered/93.1.77

Vos P, Hogers R, Bleeker M, Reijans M, Lee Tvd, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23 (21):4407-4414. doi:10.1093/nar/23.21.4407

Wang YX, Zhang B, Sun XY, Tan BY, Xu LA, Huang MR, Wang MX (2011) Comparative genome mapping among Populus adenopoda, P. alba, P. deltoides, P. euramericana and P. trichocarpa. Genes Genet Syst 86 (4):257-268

Wang ZJ, Zhu JY, Zalesny RS, Chen KF (2012) Ethanol production from poplar wood through enzymatic saccharification and fermentation by dilute acid and SPORL pretreatments. Fuel 95 (1):606-614. doi:10.1016/j.fuel.2011.12.032

Wegrzyn JL, Eckert AJ, Choi MY, Lee JM, Stanton BJ, Sykes R, Davis MF, Tsai CJ, Neale DB (2010) Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (Populus trichocarpa, Salicaceae) secondary xylem. New Phytologist 188 (2):515-532. doi:10.1111/j.1469-8137.2010.03415.x

Weih M, Bonosi L (2009) Assessment of Genotype Ranking in Long-term Biomass Production of Salix Based on Juvenile Plant Traits: Breeding Implications. Bioenergy Research 2 (1-2):29-36. doi:10.1007/s12155-009-9031-4

Weih M, Bonosi L, Ghelardini L, Ronnberg-Wastljung AC (2011) Optimizing nitrogen economy under drought: increased leaf nitrogen is an acclimation to water stress in willow (Salix spp.). Annals of Botany 108 (7):1347-1353. doi:10.1093/aob/mcr227

Weih M, Ronnberg-Wastljung AC, Glynn C (2006) Genetic basis of phenotypic correlations among growth traits in hybrid willow (Salix dasyclados x S-viminalis) grown under two water regimes. New Phytologist 170 (3):467-477. doi:10.1111/j.1469-8137.2006.01685.x

Wessels CB, Malan FS, Rypstra T (2011) A review of measurement methods used on standing trees for the prediction of some mechanical properties of timber. European Journal of Forest Research 130 (6):881-893. doi:10.1007/s10342-011-0484-6

Wilkinson JM, Evans EJ, Bilsborrow PE, Wright C, Hewison WO, Pilbeam DJ (2007) Yield of willow cultivars at different planting densities in a commercial short rotation coppice in the north of England. Biomass Bioenerg 31 (7):469-474. doi:10.1016/j.biombioe.2007.01.020

Woolbright SA, DiFazio SP, Yin T, Martinsen GD, Zhang X, Allan GJ, Whitham TG, Keim P (2008) A dense linkage map of hybrid cottonwood (Populus fremontii x P-angustifolia) contributes to long-term ecological research and comparison mapping in a model forest tree. Heredity 100 (1):59-70. doi:10.1038/sj.hdy.6801063

Wu R, Bradshaw HD, Stettler RF (1998) Developmental quantitative genetics of growth in Populus. Theoretical and Applied Genetics 97 (7):1110-1119. doi:10.1007/s001220050998

Wu R, Stettler RF (1997) Quantitative genetics of growth and development in Populus .2. The partitioning of genotype x environment interaction in stem growth. Heredity 78:124-134. doi:10.1038/hdy.1997.20

Wu R, Stettler RF (1998) Quantitative genetics of growth and development in Populus. III. Phenotypic plasticity of crown structure and function. Heredity 81:299-310. doi:10.1038/sj.hdy.6883970

Wu RL, Han YF, Hu JJ, Fang JJ, Li L, Li ML, Zeng ZB (2000) An integrated genetic map of Populus deltoides based on amplified fragment length polymorphisms. Theoretical and Applied Genetics 100 (8):1249-1256. doi:10.1007/s001220051431

Wu RL, Ma CX, Zhu J, Casella G (2002) Mapping epigenetic quantitative trait loci (QTL) altering a developmental trajectory. Genome 45 (1):28-33. doi:10.1139/g01-118

Wu S, Yang J, Huang Y, Li Y, Yin T, Wullschleger SD, Tuskan GA, Wu R (2010) An improved approach for mapping quantitative trait Loci in a pseudo-testcross: revisiting a poplar mapping study. Bioinformatics and biology insights 4:1-8

Wullschleger S, Yin TM, DiFazio SP, Tschaplinski TJ, Gunter LE, Davis MF, Tuskan GA (2005) Phenotypic variation in growth and biomass distribution for two advanced-generation pedigrees of hybrid poplar. Can J For Res-Rev Can Rech For 35 (8):1779-1789. doi:10.1139/x05-101

Xhaard C, Barres B, Andrieux A, Bousset L, Halkett F, Frey P (2012) Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. Molecular Ecology 21 (10):2383-2398. doi:10.1111/j.1365-294X.2012.05556.x

Xhaard C, Fabre B, Andrieux A, Gladieux P, Barres B, Frey P, Halkett F (2011) The genetic structure of the plant pathogenic fungus Melampsora larici-populina on its wild host is extensively impacted by host domestication. Molecular Ecology 20 (13):2739-2755. doi:10.1111/j.1365-294X.2011.05138.x

Yang XH, Kalluri UC, DiFazio SP, Wullschleger SD, Tschaplinski TJ, Cheng ZM, Tuskan GA (2009) Poplar Genomics: State of the Science. Critical Reviews in Plant Sciences 28 (5):285-308. doi:10.1080/07352680903241014

Yeh TF, Chang HM, Kadla JF (2004) Rapid prediction of solid wood lignin content using transmittance near-infrared spectroscopy. Journal of Agricultural and Food Chemistry 52:1435-1439

Yin T, Zhang X, Gunter L, Priya R, Sykes R, Davis M, Wullschleger SD, Tuskan GA (2010) Differential Detection of Genetic Loci Underlying Stem and Root Lignin Content in Populus. PLoS ONE 5 (11). doi:10.1371/journal.pone.0014021

Yin TM, DiFazio SP, Gunter LE, Jawdy SS, Boerjan W, Tuskan GA (2004a) Genetic and physical mapping of Melampsora rust resistance genes in Populus and characterization of linkage disequilibrium and flanking genomic sequence. New Phytologist 164 (1):105-105. doi:10.1111/j.1469-8137.2004.01161.x

Yin TM, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA (2004b) Large-scale heterospecific segregation distortion in Populus revealed by a dense genetic map. Theoretical and Applied Genetics 109 (3):451-463. doi:10.1007/s00122-004-1653-5

Yin TM, Huang MR, Wang MX, Zhu LH, Zeng ZB, Wu RL (2001) Preliminary interspecific genetic maps of the Populus genome constructed from RAPD markers. Genome 44 (4):602-609. doi:10.1139/gen-44-4-602

Yordanov YS, Regan S, Busov V (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. doi:10.1105/tpc.110.078634

Zalesny RS, Bauer EO (2007) Evaluation of Populus and Salix continuously irrigated with landfill leachate II. Soils and early tree development. International Journal of Phytoremediation 9 (4):307-323. doi:10.1080/15226510701476594

Zamudio FJ, Yanez M (2010) Review of the research conducted by the Poplar Technology Center. Open Conference System, Fifth International Poplar Symposium (IPS V)

Zhang BH, Wang QL, Pan XP (2007) MicroRNAs and their regulatory roles in animals and plants. Journal of Cellular Physiology 210 (2):279-289. doi:10.1002/jcp.20869

Zhang D, Zhang Z, Yang K (2006) QTL analysis of growth and wood chemical content traits in an interspecific backcross family of white poplar (Populus tomentosa x P-bolleana) x P-tomentosa. Can J For Res-Rev Can Rech For 36 (8):2015-2023. doi:10.1139/x06-103

Zhang D, Zhang Z, Yang K, Li B (2004) Genetic mapping in (Populus tomentosa x Populus bolleana) and P-tomentosa Carr. using AFLP markers. Theoretical and Applied Genetics 108 (4):657-662. doi:10.1007/s00122-003-1478-7

Zhang J, Steenackers M, Storme V, Neyrinck S, Van Montagu M, Gerats T, Boerjan W (2001) Fine mapping and identification of nucleotide binding site/leucine-rich repeat sequences at the MER locus in Populus deltoides 'S9-2'. Phytopathology 91 (11):1069-1073. doi:10.1094/phyto.2001.91.11.1069

Zhang ZH, Yu JY, Li DF, Zhang ZY, Liu FX, Zhou X, Wang T, Ling Y, Su Z (2010) PMRD: plant microRNA database. Nucleic Acids Research 38:D806-D813. doi:10.1093/nar/gkp818

Zhao JY, Dimov Z, Becker HC, Ecke WG, Mollers C (2008) Mapping QTL controlling fatty acid composition in a doubled haploid rapeseed population segregating for oil content. Mol Breed 21 (1):115-125. doi:10.1007/s11032-007-9113-y

Zhong R, Ye Z-H (2009) Transcriptional regulation of lignin biosynthesis. Plant Signal Behav 4 (11):1028-1034

Zhong R, Ye Z-H (2010) The poplar PtrWNDs are transcriptional activators of secondary cell wall biosynthesis. Plant Signal Behav 5 (4):469-472

Zhong RQ, Lee CH, Ye ZH (2010) Functional characterization of poplar wood-associated NAC domain transcription factors. Plant Physiology 152 (2):1044-1055. doi:10.1104/pp.109.148270

Zhou FF, Xu Y (2009) RepPop: a database for repetitive elements in Populus trichocarpa. BMC Genomics 10 (14). doi:10.1186/1471-2164-10-14

Zhou GW, Taylor G, Polle A (2011) FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. Plant Methods 7. doi:10.1186/1746-4811-7-9

Zhou JL, Lee CH, Zhong RQ, Ye ZH (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. doi:10.1105/tpc.108.063321

Zhu JY, Pan X, Zalesny RS, Jr. (2010) Pretreatment of woody biomass for biofuel production: energy efficiency, technologies, and recalcitrance. Applied Microbiology and Biotechnology 87 (3):847-857. doi:10.1007/s00253-010-2654-8

Zhu Q-H, Guo A-Y, Gao G, Zhong Y-F, Xu M, Huang M, Luo J (2007) DPTF: a database of poplar transcription factors. Bioinformatics 23 (10):1307-1308. doi:10.1093/bioinformatics/btm113

Zsuffa L (1974) The genetics of Populus nigra L. Annales Forestales 6 (2):29-53

Redouane El Malki

Architecture génétique des caractères cibles pour la culture du peuplier en taillis à courte rotation

Résumé :

L'optimisation de la biomasse produite par les taillis à courte rotation de peupliers représente un enjeu majeur pour la production de biocarburants de deuxième génération. Dans ce contexte, ce travail vise à faciliter le développement à court terme de nouvelles variétés clonales de peuplier permettant la production d'une ressource de qualité en s'intéressant plus particulièrement à l'architecture génétique de la résistance à la rouille foliaire et de la qualité du bois chez le peuplier noir (*Populus nigra*), espèce parente des hybrides cultivés. Des marqueurs SNP ont été développés à partir du séquençage de 665 fragments de gènes dans un panel de 21 individus. Ces derniers ont été associés à des marqueurs SSR et AFLP pour construire de nouvelles cartes génétiques sur une famille de 324 plein-frères clonés. Une technique de phénotypage à haut débit basée sur la spectrométrie à proche infrarouge a été développée pour prédire les teneurs en composés chimiques du bois ainsi que le rendement en saccharification. La mise en évidence d'une variabilité génétique importante pour l'ensemble des caractères a permis de cartographier les régions génomiques impliquées dans leur variation. Parmi les 11 QTL détectés pour la résistance, un QTL à effet majeur co-localise avec un QTL majeur associé à la résistance à la rouille foliaire du saule. Pour la qualité du bois, 15 QTL à effet faible à moyen ont été détectés, dont un cartographié sur le chromosome XIII qui co-localise avec des QTL précédemment identifiés chez le peuplier pour les teneurs en sucres et en lignines. Ce travail de thèse ouvre des perspectives d'identification de gènes sous-jacents aux QTL par génétique d'association.

Mots clés : peuplier noir, bioéthanol, résistance à la rouille foliaire, qualité du bois, diversité nucléotidique, cartographie génétique, variabilité génétique, QTL

Genetic architecture of target traits for short rotation coppice poplar

Abstract :

Improvement of lignocellulosic resources from poplar short rotation coppices is a major challenge for the production of second generation biofuels. In this context, the present work aims at optimizing short term creation and deployment of improved poplar clonal varieties through the dissection of genetic control of both leaf rust resistance and wood quality in black poplar (*Populus nigra*), one of the parental species of cultivated hybrids. SNP markers have been developed from the resequencing of 665 gene fragments in a discovery panel of 21 individuals. These markers were combined with SSRs and AFLPs to build new genetic maps in a pedigree composed of 324 cloned full-sibs. High throughput phenotyping based on near infrared spectroscopy has been used to predict wood chemical contents and saccharification yield. High genetic variability expressed in all traits allowed the identification of genomic regions controlling this variation. Of the 11 QTL mapped for resistance, one major QTL co-localized with a major QTL previously detected for leaf rust resistance in willow. For wood quality, 15 QTL with low to moderate effect have been identified. Interestingly, one QTL mapped on chromosome XIII and colocalized with sugar and lignin contents QTL previously detected in poplar. Present results open perspectives towards the identification of candidate genes underlying the detected QTL through association genetics.

Keywords : black poplar, bioethanol, leaf rust resistance, wood quality, nucleotide diversity, genetic mapping, genetic variability, QTL

UR0588, Amélioration Génétique et Physiologie Forestières, INRA Orléans