



# Optimisation des stratégies d'amélioration génétique du pin maritime grâce à l'utilisation de marqueurs moléculaires

Marjorie Vidal

► **To cite this version:**

Marjorie Vidal. Optimisation des stratégies d'amélioration génétique du pin maritime grâce à l'utilisation de marqueurs moléculaires. *Ecologie, Environnement*. Université de Bordeaux, 2016. Français. <NNT : 2016BORD0044>. <tel-01359191>

**HAL Id: tel-01359191**

**<https://tel.archives-ouvertes.fr/tel-01359191>**

Submitted on 2 Sep 2016

**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.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

**DOCTEUR DE  
L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE SCIENCES ET ENVIRONNEMENTS  
SPÉCIALITÉ Ecologie évolutive, fonctionnelle et des communautés

Par Marjorie VIDAL

**Optimisation des stratégies d'amélioration génétique  
du pin maritime grâce à l'utilisation de marqueurs  
moléculaires**

Sous la direction de : Laurent BOUFFIER, Christophe PLOMION et Luc HARVENGT.

Soutenue le 06 avril 2016.

Membres du jury :

M. GOMES, Éric	Professeur, Université de Bordeaux, France	Président du jury
M. MACKAY, John	Professeur, Université d'Oxford, Angleterre	Rapporteur
M. GENGLER, Nicolas	Professeur, Université de Liège - Gembloux, Belgique	Rapporteur
M. SANCHEZ, Leopoldo	Chargé de Recherche, INRA - Centre Val de Loire, France	Examinateur
Mme VITEZICA, Zulma	Maître de Conférences, ENSAT- INP Toulouse, France	Examinatrice
M. PLOMION, Christophe	Directeur de Recherche, INRA Bordeaux, France	Directeur de thèse
M. BOUFFIER, Laurent	Chargé de Recherche, INRA Bordeaux, France	Co-encadrant
M. HARVENGT, Luc	Responsable Scientifique, FCBA, Bordeaux, France	Co-encadrant



## Titre : Optimisation des stratégies d'amélioration génétique du pin maritime grâce à l'utilisation de marqueurs moléculaires

### Résumé

Le pin maritime (*Pinus pinaster* Ait.) est l'une des principales espèces forestières en France, fournissant près d'un quart de la production nationale de bois. Un programme d'amélioration, mis en place dans les années 1960, propose des variétés génétiquement améliorées pour la croissance et la rectitude du tronc.

Cette thèse explore la possibilité d'introduire les marqueurs moléculaires dans les stratégies d'amélioration génétique du pin maritime en Aquitaine. Les marqueurs sont utilisés afin de reconstituer *a posteriori* les pedigrees au sein d'un test de descendance « polycross », pour d'une part vérifier les hypothèses sur lesquelles repose la sélection *backward*, et d'autre part, pour proposer une stratégie de sélection innovante. Tout d'abord, la reconstitution du pedigree de 984 individus à l'aide de 63 marqueurs SNPs permet de valider les hypothèses de la sélection *backward*, et montre que les estimations des paramètres génétiques et des valeurs génétiques maternelles, basées sur l'information d'un pedigree partiel ou complet, diffèrent peu. Puis, les meilleurs descendants du test polycross sont présélectionnés et génotypés pour évaluer la faisabilité d'une stratégie de sélection *forward*. Enfin, des vergers à graines sont simulés selon différentes stratégies de sélection (*backward*, *forward*, mixte) afin de comparer les gains génétiques des variétés améliorées ainsi obtenues.

Une stratégie de sélection *forward* chez le pin maritime permettrait d'accélérer les cycles de sélection et d'augmenter la fréquence des sorties variétales. De plus, le jeu de marqueurs SNPs développé dans cette étude est en cours de valorisation dans différentes étapes du programme d'amélioration.

**Mots clés :** *Pinus pinaster*, amélioration génétique, reconstitution de pedigree, marqueurs SNP, test polycross, stratégies de sélection, gain génétique.

## Title: Optimization of maritime pine breeding strategies using molecular markers

### Abstract

Maritime pine (*Pinus pinaster* Ait.) is one of the main economical forest species in France, providing about twenty five percent of the national round wood production. A breeding program, implemented since the 60's, offers genetically improved varieties for growth and stem straightness.

This PhD explores the use of molecular markers in breeding strategies for maritime pine in Aquitaine. Molecular markers were used for pedigree recovery in a polycross progeny trial to test assumptions of backward selection on one hand, and to evaluate the feasibility of a new breeding strategy on the other hand. First, the pedigree of 984 progeny was recovered with 63 SNPs allowing to verify the assumptions of backward selection. We also showed that genetic parameters and maternal breeding value estimates were not much modified by inclusion of full pedigree information. Then, the best progenies in the polycross trial were preselected and genotyped to investigate the possibility of carrying out a forward selection strategy. Finally, establishment of clonal seed orchards were simulated from various breeding strategies (backward, forward, mixed) in order to compare genetic gains from the improved varieties obtained thereby.

This study opens new perspectives towards an implementation of forward selection in the French maritime pine breeding program, to speed the selection cycles up and to increase the frequency of variety renewal. Moreover, the set of SNP markers developed is now used in different steps of the breeding program.

**Keywords:** *Pinus pinaster*, genetic improvement, pedigree recovery, SNP markers, polycross trial, breeding strategies, genetic gain.



## *Avant-propos*

Ce travail a été effectué dans le cadre d'une thèse CIFRE (n°947/2012), au sein du pôle Biotechnologies et Sylviculture Appliquée de l'Institut Technologique FCBA (Forêt Cellulose Bois-construction Ameublement) et de l'UMR Biodiversité Gènes et Communautés (BioGeCo) à l'INRA (Institut National de la Recherche Agronomique) de Cestas Pierroton.

Ces travaux de recherche ont été financés sur une durée de 3 ans par FCBA, notamment au travers du projet Européen ProCoGen (<http://www.procogen.eu/>) et du projet régional IMAF (CR Aquitaine), et par l'Association Nationale de la Recherche et de la Technologie. Ils s'intègrent dans le cadre du programme d'amélioration génétique du pin maritime, géré par le groupement d'intérêt scientifique « Pin Maritime du Futur ».

Un séjour de 4 mois (août-décembre 2014) à la *North Carolina State University* (Raleigh, Caroline du Nord, Etats-Unis) au sein de la coopérative *Tree Improvement Program* a été financé par le LabEx COTE et l'IdEx de l'Université de Bordeaux.

### **Laboratoires d'accueil**

**INRA** - Institut National de la Recherche  
Agronomique  
UMR 1202 – Biodiversité, Gènes et Ecosystèmes  
Equipe Ecologie et Génomique Fonctionnelle  
69, route d'Arcachon 33612 CESTAS - France



**FCBA** - Institut Technologique Forêt Cellulose Bois-  
construction Ameublement  
Pôle Biotechnologie et Sylviculture avancée  
Equipe Génétique et Biotechnologies  
71, route d' Arcachon 33610 CESTAS - France



### **Encadrement**

Directeur de thèse : **Christophe Plomion**, Directeur de recherche, HDR, INRA

Co-encadrants : **Laurent Bouffier**, Chargé de recherche, INRA

**Luc Harvengt**, Responsable Scientifique Biotechnologie, FCBA

### **Dates de la thèse**

Janvier 2013 – Janvier 2016

## Remerciements

*En démarrant ma thèse début 2013, je ne m'attendais pas à vivre une telle aventure. Maintenant qu'elle touche à sa fin, c'est avec une grande émotion que je souhaite remercier l'ensemble des personnes qui y ont participé et en ont fait une belle aventure, enrichissante tant sur le plan professionnel que personnel.*

*Tout d'abord, je tiens à remercier très chaleureusement Laurent Bouffier, qui m'a encadré et guidé tout au long de ces 3 années. Un grand merci pour tes précieux conseils, ta patience et ta confiance. Ça a été un réel plaisir de travailler à tes côtés. Merci aussi à mon directeur de thèse Christophe Plomion, et à mon tuteur scientifique Luc Harvengt, pour leurs conseils et leurs relectures constructives de mes deux articles.*

*J'exprime également mes sincères remerciements aux membres du jury pour avoir accepté d'évaluer ce travail et pour leur participation à ma soutenance : John Mackay, Nicolas Gengler, Éric Gomes, Leopoldo Sanchez, et Zulma Vitezica.*

*Merci à l'institut technologique FCBA et à l'Association Nationale de la Recherche et de la Technologie pour le financement de cette thèse, mais aussi à l'UMR Biogeco de l'INRA pour les nombreuses formations.*

*Merci aussi à tous les membres de mes comités de thèse pour leurs conseils et les discussions constructives.*

*Un très grand merci à Annie Raffin pour son aide tout au long de cette thèse, sa gentillesse, ses précieux conseils et ses relectures.*

*Merci à Jérôme Bartholomé pour son aide à dompter R, ses relectures constructives, et pour toutes nos discussions sur la sélection génomique.*

*Merci à Patrick Léger, Christophe Boury, Adline Delcamp et Franck Salin pour l'aide au labo et pour le génotypage.*

*Merci également à Sophie Gerber, Isabelle Lesur, François Erhenman pour les coups de main « statistiques » ou « bio-informatiques ».*

*Merci à Thierry Labbé et Loïc Kerdraon pour le soutien informatique, et à Florence Le Pierres, Véronique Lozano et Patricia Poupeau pour le soutien administratif.*

*Merci à toutes les personnes qui m'ont accordé un peu de leur temps pour relire ce manuscrit : Priscilla Cailly, Annie Raffin, Céline Méredieu, Jérôme Bartholomé, et Jocelyn Domange.*

*Merci à Fikret Isik pour son invitation et son accueil à NCSU. Merci à toute l'équipe du Tree Improvement Program pour m'avoir fait découvrir l'amélioration du taeda et les joies de la vie d'un campus américain. Merci également à l'IdEx, au LabEx et à FCBA pour avoir financé ce séjour de 4 mois à NCSU. Merci à Alain Bailly pour sa confiance et pour m'avoir laissée partir ces quelques mois outre-Atlantique.*

*Merci à tous ceux grâce à qui j'ai passé de très bons moments tout au long de ces 3 ans. Merci à tous les « occupants » de l'Artiga pour la bonne humeur quotidienne : Hélène, Adline, Laura, Mélanie, Thibaut, Isa, Céline, Emilie, Christophe, Adib, Greg, Erwan, ... Bon courage à Thomas, Gaby, Adib, Nastasia pour votre fin de thèse.*

*Merci également à l'ensemble des membres de la station Sud-Ouest de FCBA pour la bonne ambiance de travail. Et plus particulièrement, merci Jean-Mathieu pour nos mythiques sorties de récolte d'aiguilles, pour ta bienveillance, ton humour (pas toujours du meilleur goût!), et pour le « BFC » ; merci Priscilla pour ton soutien sans faille les derniers mois (« je te l'avais dit ! ») ; merci Karine pour tes encouragements, ton soutien et pour les pauses cafés ; merci Ambre pour nos ptits moments de folies ; merci Patricia pour l'aide dans la gestion administrative ; merci Jean-Pierre et Pierre pour votre aide sur le travail de terrain.*

*Enfin, un grand merci à mes amis et à ma famille (avec une mention spéciale à ma maman) pour leurs encouragements tout au long de ce travail.*



# Table des matières

<b>Abréviations</b> .....	<b>10</b>
<b>1. Introduction</b> .....	<b>11</b>
<b>1.1 Les forêts de plantation, un atout pour répondre à la demande en bois</b> .....	<b>13</b>
1.1.1 A l'échelle mondiale : une forte demande en bois, satisfaite en partie grâce aux forêts plantées .....	13
1.1.2 Le massif des Landes de Gascogne, plus grande forêt plantée française .....	15
1.1.3 La forêt plantée landaise face à de nouveaux défis .....	19
<b>1.2 L'amélioration génétique du pin maritime en Aquitaine</b> .....	<b>23</b>
1.2.1 Objectifs et contraintes du programme d'amélioration du pin maritime .....	23
1.2.2 Une stratégie de sélection récurrente.....	25
1.2.3 Constitution des populations d'amélioration successives .....	29
1.2.4 Evaluation génétique des individus de la population d'amélioration.....	30
1.2.5 Déploiement du gain génétique des variétés améliorées.....	37
1.2.6 Gestion de la diversité génétique.....	39
<b>1.3 Intégration des marqueurs moléculaires dans les programmes d'amélioration     forestiers</b> .....	<b>47</b>
1.3.1 Définitions et utilisations.....	47
1.3.2 Des outils pour prédire le phénotype.....	48
1.3.3 Des outils d'aide dans les programmes d'amélioration génétique dès aujourd'hui, grâce à la reconstitution de pedigree .....	51
<b>2. Problématique et objectifs de la thèse</b> .....	<b>55</b>
<b>3. Matériel et méthodes</b> .....	<b>59</b>
<b>3.1 Test de descendance polycross étudié</b> .....	<b>59</b>
<b>3.2 Phénotypage des caractères d'intérêt</b> .....	<b>59</b>
<b>3.3 Echantillonnage des descendants du test polycross</b> .....	<b>61</b>
<b>3.4 Génotypage</b> .....	<b>61</b>
3.4.1 Conception du jeu de SNPs.....	61
3.4.2 Génotypage iPLEX® Gold (Sequenom).....	63
<b>3.5 Reconstitution de pedigree : méthode</b> .....	<b>63</b>
<b>3.6 Evaluation génétique</b> .....	<b>64</b>
<b>4. Reconstitution de pedigree dans un test de descendance polycross et conséquences pour la stratégie de sélection <i>backward</i></b> .....	<b>65</b>
<b>4.1 Résumé de l'article</b> .....	<b>65</b>
<b>4.2 Article 1</b> .....	<b>66</b>
<b>4.3 Conclusion de la partie 4</b> .....	<b>80</b>

<b>5. Evaluation d'une stratégie de sélection <i>forward</i> de type « <i>polymix breeding</i> » .....</b>	<b>81</b>
<b>5.1 Résumé de l'article .....</b>	<b>81</b>
<b>5.2 Article 2.....</b>	<b>82</b>
<b>5.3 Conclusion de la partie 5.....</b>	<b>103</b>
<b>6. Applications et perspectives.....</b>	<b>105</b>
<b>6.1 Les premières utilisations des marqueurs moléculaires dans le programme d'amélioration du pin maritime.....</b>	<b>105</b>
6.1.1 Utilisation du jeu de SNPs développé pendant la thèse.....	105
6.1.2 Des SNPs pour certifier l'origine des matériels de reboisement.....	107
<b>6.2 Vers la valorisation des tests de descendance polycross existants ?.....</b>	<b>109</b>
<b>6.3 Vers la mise en place d'une stratégie de sélection forward dans le programme d'amélioration .....</b>	<b>111</b>
<b>6.4 Vers la sélection génomique .....</b>	<b>112</b>
<b>7. Conclusion générale.....</b>	<b>117</b>
<b>Bibliographie.....</b>	<b>119</b>
<b>Annexe 1: Liste des publications et communications.....</b>	<b>129</b>

## Abréviations

ADN : Acide désoxyribonucléique  
BLUE : Best Linear Unbiased Estimators  
BLUP : Best Linear Unbiased Predictors  
CPFA : Centre de Productivité Forestière d'Aquitaine  
CRPF : Centre Régional de la Propriété Forestière  
DRAAF: Direction Régionale de l'Alimentation, de l'Agriculture et de la Forêt  
EBV : Estimated breeding value  
FAO : Food and Agriculture Organization  
FCBA : institut technologique Forêt Cellulose Bois-construction Ameublement  
GEBV : Genomic estimated breeding value  
GIEC : Groupe d'experts intergouvernemental sur l'évolution du climat  
GIS PMF : groupement d'intérêt scientifique Pin Maritime du Futur  
IBD : Identity by descent  
IGN : Institut national de l'information géographique et forestière  
INRA : Institut National de la Recherche Agronomique  
IUFRO : International Union of Forest Research Organizations  
LOD : Likelihood Odds Ratio  
MAF : Minor Allele Frequency  
MALDI-TOF : Matrix-Assisted Laser Desorption/Ionisation-Time of Flight  
ONF : Office National des Forêts  
PCR : Polymerase Chain Reaction  
PMX : polymix (mélange de pollen)  
QTL : Quantitative Trait Locus  
REML : Restricted maximum likelihood  
SAM : Sélection Assistée par Marqueurs  
SBE : Single base-extensions  
SNP : Single nucleotide polymorphism  
SSR : Simple Sequence Repeat  
UMR : Unité mixte de recherche  
VG : Verger à Graines

## **1. Introduction**

La filière forêt-bois doit faire face à une forte croissance de la demande en produits forestiers, dans un contexte mondial en constante évolution : diversification des usages du bois en lien avec la raréfaction des ressources fossiles, attente sociétale pour une gestion durable des forêts, mais aussi interrogation sur les capacités d'adaptation des forêts face au changement climatique. La surface forestière mondiale diminue et les besoins mondiaux en bois s'accroissent, la pression de prélèvement exercée sur les forêts primaires et naturelles augmente alors qu'il est souhaitable de les préserver pour des raisons socio-environnementales. Il est donc nécessaire d'augmenter la productivité de certaines forêts, gérées de façon durable et localisées dans des zones accessibles pour optimiser la récolte de bois. Les forêts de plantation, basées sur la régénération artificielle (par semis ou plantation de variétés) combinée à une sylviculture adaptée, permettent un accroissement majeur de la production. Ainsi, les forêts de plantation constituent un levier essentiel pour produire du bois de qualité en grande quantité et dans des lieux accessibles. La production et la superficie des forêts de plantation dans le monde sont en constante augmentation, ce qui permet d'atténuer la pression de prélèvement sur les autres forêts. Les variétés utilisées pour les plantations sont généralement issues de la reproduction sélective d'individus ayant des caractéristiques désirables dans le cadre de programmes d'amélioration génétique. Le processus d'amélioration comprend un cycle continu de sélection, d'évaluation et de croisements, permettant d'améliorer la qualité génétique d'une essence. L'application des principes de génétique quantitative permet d'améliorer la croissance et/ou la qualité des arbres utilisés en plantation. Bien que l'augmentation du volume de la tige soit généralement l'objectif prioritaire, les améliorations relatives à la résistance aux maladies, à la rectitude du tronc et à la qualité du bois et des fibres peuvent aussi avoir une grande importance selon les essences et leurs usages. De plus, ce processus d'amélioration génétique s'inscrit dans un contexte environnemental changeant : les variétés doivent être adaptées aux nouvelles contraintes imposées par le changement climatique. Ainsi, l'amélioration génétique forestière joue un rôle crucial pour maintenir et même augmenter les approvisionnements en bois. Elle doit fournir des variétés productives adaptées aux différentes attentes de la société (économiques, sociales et environnementales) mais aussi aux contraintes bioclimatiques.

Le massif forestier des Landes de Gascogne, situé dans le sud-ouest de la France (Aquitaine), est la plus grande forêt de plantation d'Europe occidentale et illustre bien les enjeux et les défis décrits précédemment. Pour améliorer la productivité de ce massif et pérenniser la ressource bois en Aquitaine, la filière forêt-bois compte sur l'optimisation des techniques sylvicoles et l'amélioration génétique du pin maritime, l'espèce majoritaire de la forêt landaise. La création de variétés améliorées constitue ainsi un enjeu stratégique majeur qui conditionne le développement durable de ce secteur. L'amélioration génétique est donc essentielle pour répondre aux objectifs de la politique nationale et régionale de la filière forêt-bois : elle permet d'adapter au mieux la production de bois aux besoins des industriels en aval, mais elle a également pour objectif d'aider le massif forestier à s'adapter au changement climatique.

Figure 1 : Superficie des forêts plantées, par domaine climatique, entre 1990 et 2015 (FAO 2015)

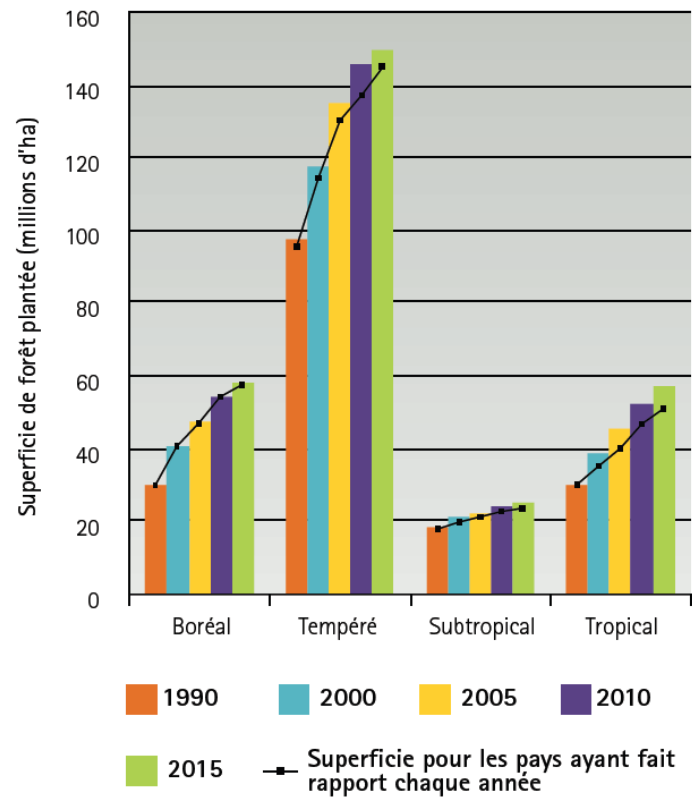
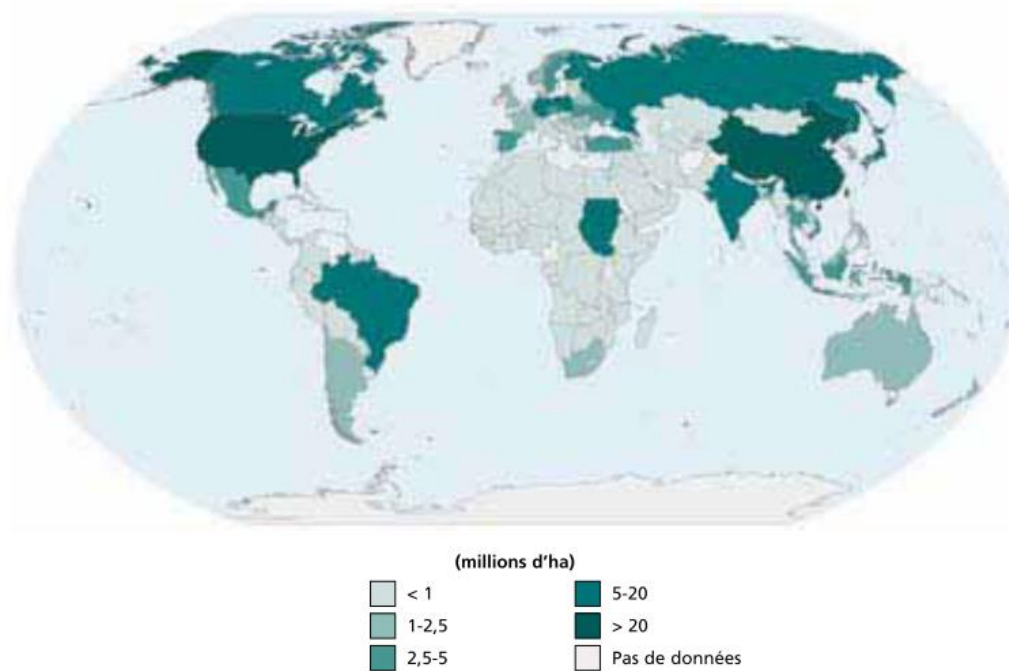


Figure 2 : Superficie des forêts plantées en 2010, par pays (FAO 2010)



Cette introduction décrit d'abord la situation et le rôle des forêts plantées dans le monde et en France, et plus particulièrement les plantations de pin maritime du massif des Landes de Gascogne. Ensuite, le programme d'amélioration du pin maritime en Aquitaine, dans lequel s'inscrit cette thèse, est présenté. Enfin, la possibilité d'utiliser des marqueurs moléculaires au sein du programme d'amélioration dans le but d'établir de nouvelles approches pour améliorer les stratégies de sélection, est discutée.

## 1.1 Les forêts de plantation, un atout pour répondre à la demande en bois

### 1.1.1 A l'échelle mondiale : une forte demande en bois, satisfaite en partie grâce aux forêts plantées

Dans un contexte de croissance démographique mondiale et de politiques environnementales incitant à augmenter la part des énergies et des matériaux renouvelables, la demande en bois ne cesse de croître et de se diversifier : pâte à papier, constructions en bois, charbon de bois, énergie de substitution (FAO 2010 ; IUFRO 2015). Les projections les plus récentes de la FAO estiment que la consommation mondiale de bois rond<sup>1</sup> industriel dépassera de 60 % les niveaux actuels d'ici 2030. La consommation de papier et de cartonnages pourrait également augmenter considérablement, étant donné le développement du e-commerce et l'augmentation du niveau de vie dans les pays émergents.

Le développement des forêts de plantation est indispensable pour subvenir à ces besoins croissants tout en limitant l'impact de la déforestation des milieux fragiles. Selon la FAO (définition du rapport FRA 2000), « *les forêts plantées (ou de plantation) comprennent les arbres établis par la plantation de semis et/ou par le semis direct. Les espèces peuvent être d'origine locale ou introduites. La mise en place peut être réalisée sur des terres autrefois boisées ou sur des terres qui ne l'ont jamais été auparavant. L'objectif de ces plantations forestières peut être soit commercial, soit la protection environnementale ou soit la réhabilitation de zones dégradées (ou une combinaison)* ». Les forêts de plantation, régénérées après chaque révolution<sup>2</sup>, peuvent de ce fait contribuer à la gestion durable des milieux forestiers dans leur ensemble.

La surface des forêts de plantation a augmenté de 110 millions d'hectares au niveau mondial ces 25 dernières années, et couvre aujourd'hui 7 % (*i.e.* 277 millions ha) de la superficie totale de forêts (Payn *et al.* 2015). Entre 1990 et 2000, la superficie des forêts de plantation a augmenté (en moyenne) de 3,6 millions d'ha par an, puis de 5,2 millions d'ha/an entre 2000 et 2010, et de 3,1 millions d'ha/an entre 2010 et 2015 en raison d'une baisse des plantations en Asie de l'Est, en Europe, en Amérique du Nord, en Asie du Sud et en Asie du Sud-Est (FAO 2015). La plus vaste superficie de forêt plantée se trouve en zone tempérée (150 millions d'ha), suivie des zones tropicale et boréale (57 millions d'ha chacune) (FAO 2015) (Figures 1 et 2).

---

<sup>1</sup> Le bois rond comprend tout bois abattu et façonné, avant la première transformation industrielle : grume (tronc coupé, ébranché et revêtu de son écorce), bille, rondin ou bûche.

<sup>2</sup> Révolution : durée de renouvellement d'un peuplement séparant deux coupes rases ou régénérations successives. (Source: glossaire et guide agroforestier)

Figure 3: Répartition de la composition des peuplements forestiers en France métropolitaine (IGN 2013)

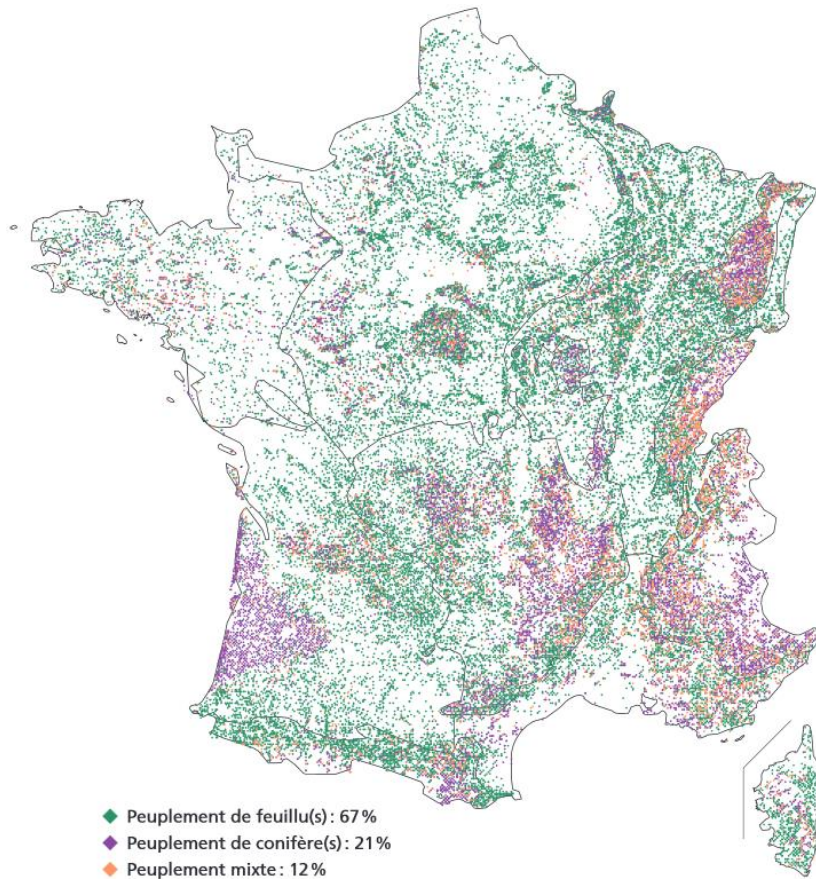
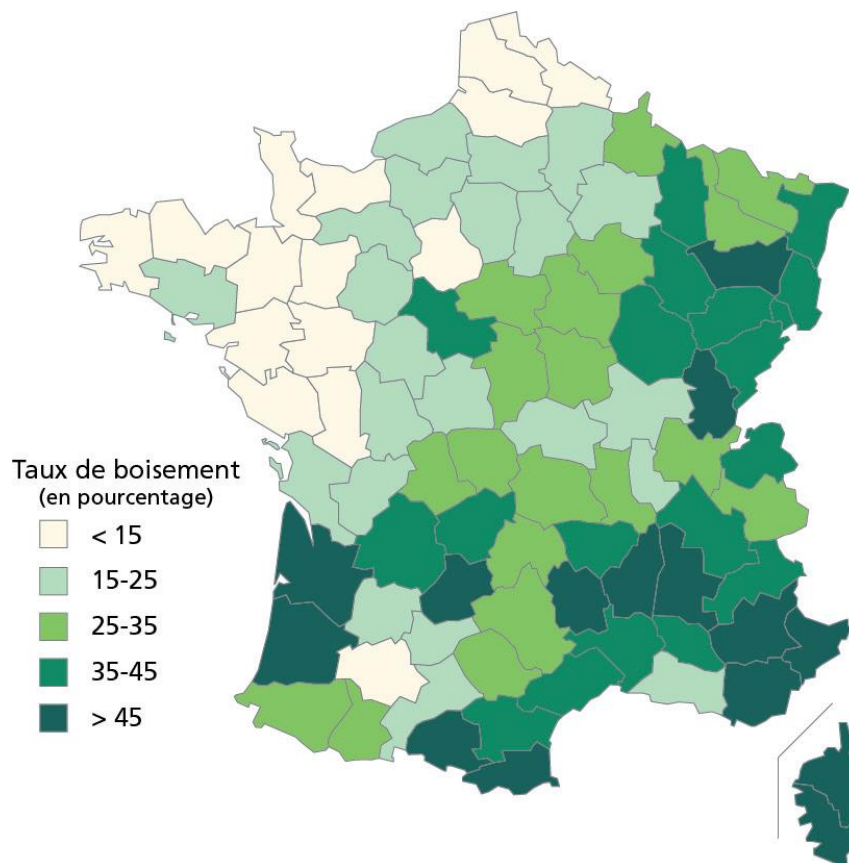


Figure 4 : Taux de boisement des départements de la France métropolitaine (IGN 2013)



Parmi les genres les plus plantés dans le monde, les conifères représentent plus de 50% des surfaces. Le genre *Pinus* totalise à lui seul 32 % des surfaces (*P. taeda*, *P. radiata*, *P. pinaster*, *P. caribea*, etc.). Les feuillus représentent 37 % des plantations avec quatre genres majeurs : *Eucalyptus* (8 %), *Acacia* (5 %), *Tectona* (4 %) et *Populus* (3 %) (FAO, States of the world's forests 2007). Les plantations tropicales sont dominées par les genres *Eucalyptus* et *Pinus* (représentant près de la moitié des superficies des plantations tropicales), et les plantations forestières tempérées et boréales sont largement dominées par le genre *Pinus* (suivi des genres *Picea* et *Abies*). Les trois quarts des forêts plantées sont constitués d'espèces indigènes, alors qu'un quart comprend des espèces introduites (FAO 2010).

Environ 75 % des forêts plantées sont destinées à la production de biomasse forestière et 25 % sont destinées à la protection environnementale (protection contre l'érosion des sols, préservation des ressources en eau, lutte contre la désertification, stabilisation des pentes, fixation des dunes de sable, fonction de brise-vent) ou à la réhabilitation de zones dégradées. En 2012, 44 % des approvisionnements mondiaux en bois rond industriels provenaient de ces forêts plantées. Les 20 premiers pays producteurs, avec en tête les Etats-Unis, le Brésil et la Chine, fournissent à eux seuls environ 675 millions m<sup>3</sup> de bois rond, soit 87 % de la production mondiale (Payn *et al.* 2015).

Le développement des forêts de plantation permettra de contribuer de plus en plus à l'approvisionnement mondial en bois, en fibre, en combustible et en produits forestiers non ligneux, tout en fournissant des services sociaux et environnementaux (Evans 2009), à une époque où la demande de ces ressources ne cesse d'augmenter. Cette évolution devrait réduire les pressions de prélèvement sur les forêts naturelles, dont la préservation est indispensable de par leur rôle écologique mais aussi socio-économique (réservoir de 80 % de la biodiversité terrestre, protection des sols contre l'érosion, lutte contre la désertification, régulation du climat et atténuation du changement climatique, préservation des ressources en eau, rôle essentiel dans le cycle du carbone, réserve alimentaire, etc.) (FAO 2010, 2014). De plus, les forêts plantées aident à compenser les effets négatifs du réchauffement de la planète, en absorbant jusqu'à 1,5 gigatonne de carbone par an, c'est-à-dire la même quantité que les émissions actuelles dues à la déforestation (Carle et Holmgren 2008).

### 1.1.2 Le massif des Landes de Gascogne, plus grande forêt plantée française

Avec 13 % de la surface boisée de l'Union Européenne, la France métropolitaine est le 3<sup>ème</sup> pays européen pour sa surface forestière après la Suède et la Finlande. La forêt française couvre plus du quart du territoire métropolitain (15,3 millions d'hectares). Elle est principalement présente dans le massif landais, sur le pourtour méditerranéen, dans les régions montagneuses et dans l'est du pays (Figure 3). La forêt française est essentiellement privée (74 %), avec 3,8 millions de propriétaires. La filière forêt-bois française emploie environ 450 000 personnes, regroupe plus de 2 600 entreprises (mobilisation, transformation ou négoce de bois) et réalise un chiffre d'affaires annuel de 60 milliards d'euros, soit près de 3% du PIB (Sénat français 2014-2015).

Cependant, il faut noter que la balance commerciale de la filière forêt-bois française est largement déficitaire, en raison de déséquilibres entre l'exportation de bois brut (les grumes notamment) et l'importation de produits transformés, comme les meubles et les papiers ou cartons (Sénat français 2014-2015).



**Figure 5 : Localisation du massif forestier des Landes de Gascogne**  
<http://www.mediaforest.net/p4-le-massif-forestier-des-landes-de-gascogne.html>



**Figure 6 : Répartition de l'aire naturelle du pin maritime (Euforgen)**  
[http://www.euforgen.org/fileadmin/templates/euforgen.org/upload/Documents/Maps/PDF/Pinus\\_pinaster.pdf](http://www.euforgen.org/fileadmin/templates/euforgen.org/upload/Documents/Maps/PDF/Pinus_pinaster.pdf)



La filière forêt-bois contribue chaque année au déficit de la balance commerciale nationale à hauteur d'environ 10 %, soit six milliards d'euros, même si la situation s'est améliorée en 2015 (Agreste 2015). En effet, le déficit a reculé de 5 % avec des importations stabilisées à 14,7 milliards d'euros et des exportations en légère hausse qui atteignent 9,2 milliards d'euros (Agreste 2015).

➤ **L'Aquitaine, première région forestière**

L'Aquitaine est la région la plus boisée de France métropolitaine (en superficie) avec un taux de boisement supérieur à 45 % (moyenne nationale : 29 %) (IGN 2013) (Figure 4). Sa forêt dédiée à la production de bois couvre 1,8 millions d'hectares, soit 11 % des surfaces boisées françaises (Agreste Aquitaine-Mémento Forêt-Bois 2015). Avec 215 millions m<sup>3</sup> de bois sur pied (DRAAF 2014), la forêt d'Aquitaine rassemble 8,5 % du volume sur pied de la forêt de production nationale (IGN 2013). Forte de cette ressource, la filière forêt-bois occupe une place significative dans l'économie régionale et attire de nombreux établissements industriels sur le territoire aquitain.

Cinq segments principaux structurent l'ensemble des activités liées à la forêt et au bois en Aquitaine : sylviculture et exploitation forestière, sciage et travail du bois, bois de construction, industrie du papier et du carton et fabrication de meubles.

➤ **Le massif des Landes de Gascogne, la plus grande forêt de plantation européenne**

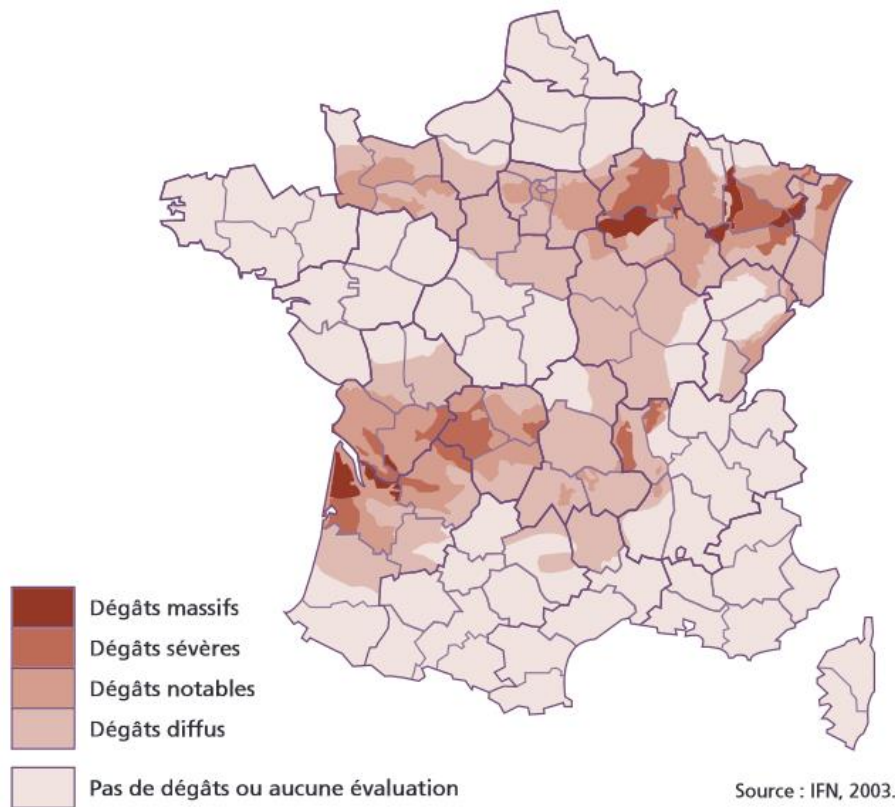
Avec près d'un million d'hectares de surface boisée, le massif forestier des Landes de Gascogne constitue la plus grande forêt cultivée d'Europe occidentale. Elle est remarquable de par son étendue et l'omniprésence du pin maritime (850 000 ha de peuplement en production en 2013). Sa bordure ouest est constituée par l'Océan Atlantique, sa bordure sud-est par l'Adour et sa bordure nord-est par la Garonne (Figure 5). Les plus grandes surfaces forestières sont concentrées dans les départements des Landes et de la Gironde, où le pin maritime occupe respectivement 84 % et 75 % de cette superficie (DRAAF 2014). Le pin maritime est donc un pilier de l'économie de la filière forêt-bois en Aquitaine, comptant 34 000 emplois directs (soit 16 % des effectifs nationaux de la filière bois), et un chiffre d'affaires généré de 2,6 milliards d'euros (FIBA, 2015).

➤ **Le pin maritime, espèce majeure du massif des Landes de Gascogne**

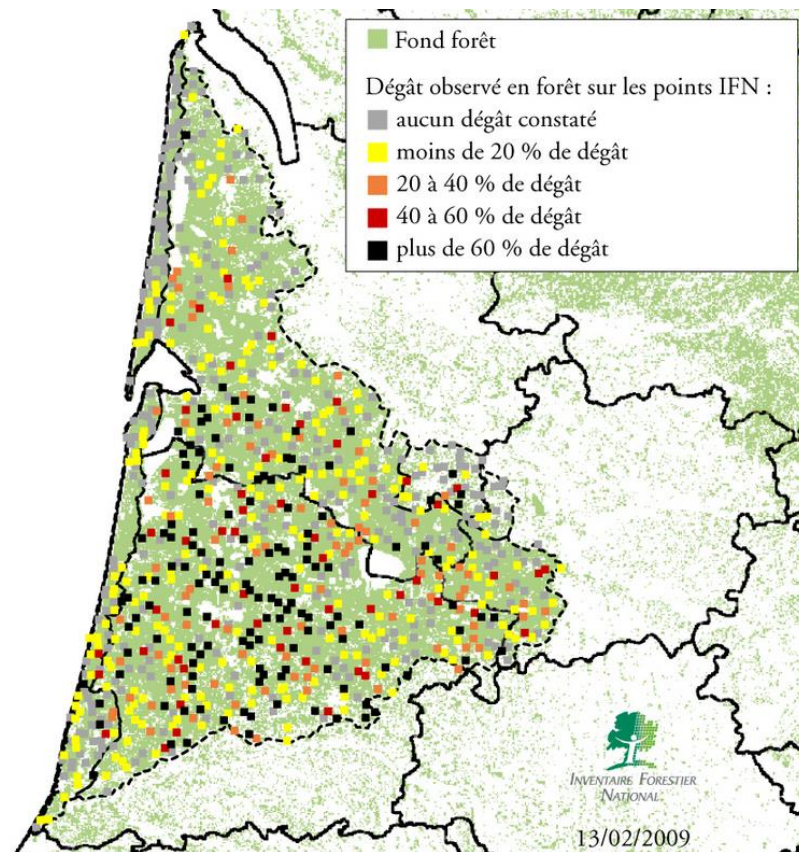
Le pin maritime, *Pinus pinaster* Aiton, est un conifère appartenant à la division des gymnospermes et à la famille des Pinacées. La superficie mondiale occupée par le pin maritime est estimée à environ 4,4 millions d'hectares, parmi lesquels 4,2 millions se situent dans sa zone d'implantation naturelle, qui s'étend du nord-ouest de la France à la Tunisie en passant par l'Espagne, le Portugal, le Maroc, la Corse et l'Italie, tout autour du bassin méditerranéen occidental (Figure 6). Cette aire naturelle de répartition couvre des conditions climatiques très variables, de 100 mm de précipitation à Oria (en Espagne) jusqu'à plus de 1000 mm à Tova (en Corse), et des environnements pédologiques variés (sable, podzols, sols rocheux). Les 200 000 hectares restants sont situés dans des zones de reforestation, où le pin maritime a été introduit avec succès (Australie, Afrique du Sud, Nouvelle-Zélande, Chili, Argentine et Uruguay).

Le pin maritime est caractérisé par une croissance rapide et une très bonne adaptation aux sols sableux pauvres et acides, qu'ils soient secs ou humides. En effet, le pin maritime tolère bien la sécheresse estivale, mais aussi l'hydromorphie (engorgement hivernal des sols).

**Figure 7 : Représentation de la sévérité des dégâts de la tempête Martin (décembre 1999) en France** (source : IFN 2003) ([http://inventaire-forestier.ign.fr/spip/IMG/pdf/L\\_IF\\_no02\\_tempetes.pdf](http://inventaire-forestier.ign.fr/spip/IMG/pdf/L_IF_no02_tempetes.pdf))



**Figure 8 : Carte de répartition des dégâts de la tempête Klaus (janvier 2009) sur le massif des Landes de Gascogne** (<http://inventaire-forestier.ign.fr/spip/IMG/jpg/AquitaineKlaus11.jpg>)



Il résiste à un froid modéré (jusqu'à -20°C pour la provenance landaise) et supporte une pluviométrie très variable. Toutes ces caractéristiques en font une espèce plastique bien adaptée aux sols aquitains (podzols majoritairement), qui sont souvent fortement acidifiés (pH 3,5 à 7), peu fertiles, gorgés d'eau en hiver et très secs en été.

Ces 50 dernières années, la production moyenne de bois est passée de 4 à près de 12m<sup>3</sup>/ha/an (CRPF 2008), essentiellement grâce aux progrès accomplis dans les techniques d'installation et de conduite des peuplements, conjugués à l'amélioration génétique. La première transformation du pin maritime est basée sur la production de bois scié (3,2 millions de m<sup>3</sup> de bois rond sur écorce consommés en 2013, soit 20 % de la production française) pour le bois d'œuvre (charpente, emballage, lambris, parquets) ainsi que l'industrie de la trituration pour les panneaux de particules et pâte de cellulose (2,5 millions de m<sup>3</sup> de bois rond sur écorce en 2013, soit 29 % de la production française) (Agreste Aquitaine-Mémento Forêt-Bois 2015).

### 1.1.3 La forêt plantée landaise face à de nouveaux défis

La filière forêt-bois en Aquitaine devra relever plusieurs défis pour assurer la pérennité de la production de bois ces prochaines années.

#### ➤ Tension sur les ressources forestières dans un contexte post-tempêtes

Après deux tempêtes successives, Martin en 1999 et Klaus en 2009, le volume de pin maritime sur pied a été réduit de moitié (140 millions de m<sup>3</sup> début 1999 contre 75 millions de m<sup>3</sup> fin 2009) (Agreste 2013). L'ouragan Martin a surtout touché le nord du massif landais (Figure 7) et a détruit 130 000 ha à plus de 50 % en Aquitaine. La tempête Klaus a plutôt touché la partie sud du massif, et a détruit près de 600 000 ha en Aquitaine, dont 223 000 ha à plus de 40 % (IFN 2009) (Figure 8). De plus, après ces deux tempêtes, des insectes xylophages (les scolytes, *Ips sexdentatus*) ont proliféré sur le bois tombé à terre, puis ont attaqué les arbres sur pied. Ce nouveau fléau a causé la perte supplémentaire de 5 millions de m<sup>3</sup> de bois après la tempête Klaus. Ces bois ont dû être exploités en urgence et étaient dévalorisés par la saturation du marché. En outre, les scolytes introduisent les agents de bleuissement (champignons du genre *Ceratostomella*) dépréciant d'autant plus le bois. Une récente étude ressource de 2013 (FCBA, IGN, INRA et CRPF Aquitaine (Thivolle-Cazat *et al.* 2013)) montre que l'approvisionnement en bois sera problématique dans les années à venir. En effet, la demande en bois ne cesse d'augmenter alors que la production du massif est ralentie dans le contexte post-tempêtes. La demande en bois d'industrie et bois énergie<sup>3</sup> atteindra au minimum 4 millions m<sup>3</sup> par an dès 2016, alors que la récolte prévue sera d'environ 2 millions m<sup>3</sup> à partir de 2018 (si l'on applique les scénarios de sylviculture actuels du pin maritime). Ce déficit en bois d'industrie de près de 2 millions m<sup>3</sup> devrait causer des problèmes d'approvisionnement et une hausse du prix du bois. Toutefois, on note que la demande en bois d'œuvre (menuiserie, ameublement, charpente, parquet) reste stable dans le temps (de l'ordre de 4 millions m<sup>3</sup> par an) et devrait être totalement satisfaite dans les prochaines années (Thivolle-Cazat *et al.* 2013).

---

<sup>3</sup> Bois énergie : utilisation du bois en tant que combustible, employé sous différentes formes - plaquettes forestières, produits connexes de scierie, granulés, bûches- dans des installations domestiques, industrielles ou collectives.

Figure 9 : Anomalies de température en Europe pour le climat de 2050 telles que simulées par le multi-modèle du GIEC (21 modèles) et suivant le scénario A1B (Source : projet Climator)

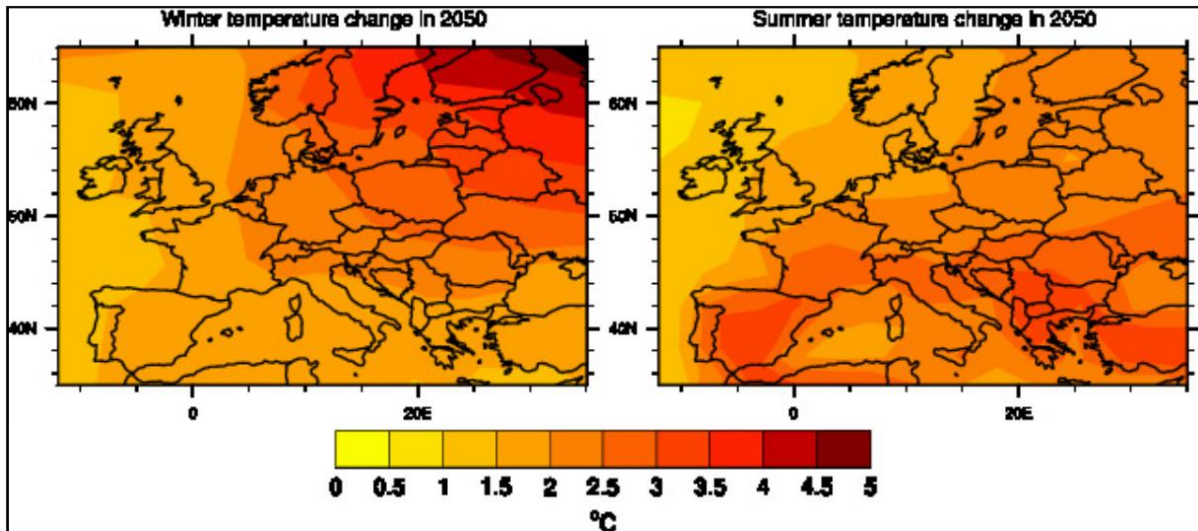


Figure 10 : Evolution du rendement du pin maritime en France : modèle GRAECO ; méthode scénario climatique AB (Source A. Bosc, 2011)

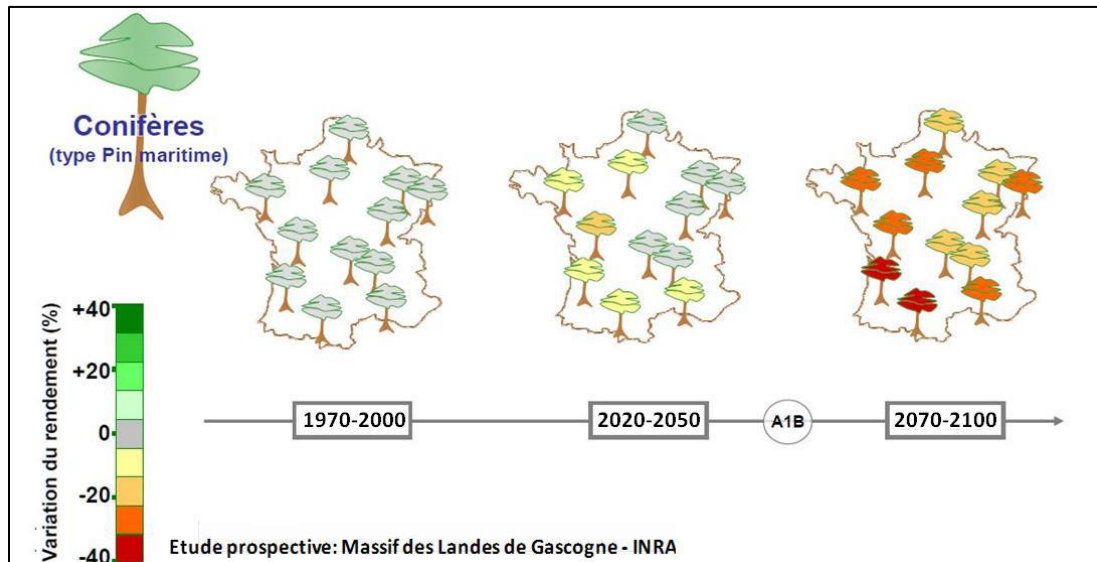
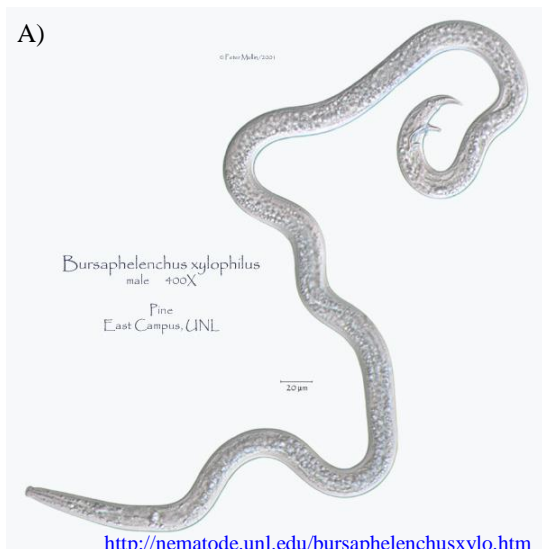


Figure 11 : A) Le nématode du pin (*Bursaphelenchus xylophilus*) et B) son insecte vecteur *Monochamus galloprovincialis*



### ➤ Impacts du changement climatique

Les études menées par le GIEC suggèrent que le réchauffement moyen en 2050 sur la France serait de l'ordre de 2°C avec une incertitude de  $\pm 0,6^\circ\text{C}$  (déterminée par les résultats de plusieurs modèles) et avec une plus forte augmentation des températures estivales, en particulier dans le sud de la France (Figure 9). Dans le sud-ouest de la France, la hausse des températures prévue par le GIEC (selon le scénario « moyen » AB1) avoisinera 1,5°C à l'horizon 2050 et 3°C à l'horizon 2100. De plus, la baisse de pluviométrie serait conséquente (de l'ordre de -100 mm à l'horizon 2050 et -200 mm à l'horizon 2100) (Terray *et al.* 2010).

Les conséquences pour l'agriculture et la forêt ont été évaluées dans le cadre du programme Climator (2010). Le changement climatique constitue une menace directe pour les écosystèmes forestiers car il a globalement un impact négatif sur la production. Bien que les écosystèmes forestiers soient, par nature, dynamiques, la rapidité des changements prévus sera probablement très supérieure à la capacité d'adaptation naturelle des écosystèmes forestiers. De plus, la majorité des forêts mondiales devrait être confrontée à l'augmentation de la fréquence des épisodes climatiques extrêmes, et notamment à des sécheresses plus intenses, ce qui entraînera une réduction de la croissance et une augmentation des agressions d'agents pathogènes (Desprez-Loustau *et al.* 2006).

En Aquitaine, la combinaison de l'augmentation des températures et de la baisse de la pluviométrie aura un impact défavorable sur la productivité du pin maritime, avec une baisse estimée entre -10 % et -20 % à l'horizon 2050, et jusqu'à -40 % après 2070 (Figure 10).

Dans ce contexte climatique, les principaux aléas naturels auxquels la forêt des Landes de Gascogne sera confrontée en 2050 vont également évoluer. D'une part, le massif landais pourrait éprouver une plus grande sensibilité aux bioagresseurs. En effet, les insectes sont généralement favorisés par une augmentation des températures. D'autre part, le massif devrait subir plus d'incendies, d'importants déficits hydriques estivaux et des engorgements hivernaux plus longs (alors que ces contraintes sont déjà marquées sur le massif Aquitain).

### ➤ Introduction du nématode

Une autre menace pour le massif Aquitain est la forte probabilité d'arrivée du nématode du pin, *Bursaphelenchus xylophilus*. Ce ver microscopique (Figure 11.A), originaire d'Amérique du Nord, aurait été introduit accidentellement au Japon au début du XX<sup>ème</sup> siècle. Puis, il s'est propagé dans d'autres pays asiatique (Chine, Taiwan, Corée) dans les années 1980 (Mallez 2014). En 1999, le nématode du pin a été détecté pour la première fois en Europe sur des pins maritimes de la région de Setubal au Portugal. De nouvelles émergences ont été signalées dans le centre du Portugal en 2008 malgré les actions gouvernementales d'éradication mises en place, puis sur l'île de Madère en 2009, ainsi qu'en Espagne en 2010 (dans les régions d'Estrémadure et de Galice).

Le nématode se développe principalement dans le bois de divers pins mais d'autres conifères ont été reconnus comme hôtes : genres *Abies*, *Cedrus*, *Larix*, *Picea*, *Pseudotsuga*. Ce nématode peut conduire à l'apparition de symptômes plus ou moins sévères, allant jusqu'à la mort des arbres dans le cas des pins européens. Le nématode du pin est véhiculé d'arbre en arbre par des coléoptères (Cerambycidae) du genre *Monochamus* (Figure 11.B). Ces insectes sont largement répandus en Aquitaine, et pourraient de ce fait disséminer rapidement le nématode dans le massif, ce qui aurait des conséquences économiques et environnementales importantes (Soliman *et al.* 2012).

➤ **Baisse de la production de graines**

Enfin, alors que la production du matériel forestier de reproduction de pin maritime repose très majoritairement sur la fructification des vergers à graines, on assiste depuis quelques années à une forte réduction de la production de graines dans ces vergers. Deux parasites principaux sont suspectés. Les acariens *Trisetacus Ehmanni* et la punaise américaine invasive *Leptoglossus occidentalis* sont considérés comme responsables d'avortements et de chutes de cônelets (inflorescences femelles) chez le pin maritime. Des études sont actuellement en cours à l'INRA et à FCBA (en liaison avec les gestionnaires des vergers et les responsables de la santé de forêts) pour identifier les ravageurs des cônes du pin maritime et analyser l'efficacité de traitements insecticides et acaricides (projets *Fortius* et *Optigraine*) sur la fructification.

Pour relever tous ces défis et assurer la pérennité de la production de bois de la forêt landaise, la production de variétés améliorées de pin maritime constitue un enjeu stratégique majeur.

## 1.2 L'amélioration génétique du pin maritime en Aquitaine

En Aquitaine, l'amélioration génétique du pin maritime débute dans les années 1960 à l'INRA et à l'AFOCEL (Association Forêt-Cellulose, devenue FCBA<sup>4</sup> depuis 2007 par fusion avec le CTBA<sup>5</sup>) avec deux programmes indépendants développés à partir d'une même population de base (arbres « plus » sélectionnés dans la forêt landaise). En 1995, ces deux instituts fusionnent leurs programmes et s'associent en créant un groupement d'intérêt scientifique, le GIS « Pin Maritime du Futur » (GIS PMF), avec pour objectif de mutualiser les moyens pour créer des variétés améliorées de pin maritime alliant productivité et qualité tout en préservant son adaptation environnementale. Trois autres organismes représentant la forêt publique et privée d'Aquitaine composent ce GIS : l'ONF (Office National des Forêts), le CPFA (Centre de Productivité Forestière d'Aquitaine) et le CRPF d'Aquitaine (Centre Régional de la Propriété Forestière Aquitaine).

Le GIS PMF concentre ses efforts d'amélioration sur la provenance landaise locale principalement, mais d'autres provenances (Corse, Espagne, Portugal, Maroc) sont également évaluées afin de développer des variétés hybrides inter-provenances. Seule l'amélioration génétique de la provenance landaise est traitée dans ce travail de thèse.

### 1.2.1 Objectifs et contraintes du programme d'amélioration du pin maritime

#### ➤ Objectifs

L'objectif du programme est d'améliorer la production de bois tant quantitativement que qualitativement. Pour assurer l'adaptation des futures variétés, un intérêt particulier est porté à la gestion de la diversité au sein des populations d'amélioration. Les deux principaux caractères retenus pour l'amélioration génétique du pin maritime sont la croissance en volume et la rectitude basale du tronc. Les gains de croissance (production totale de bois en volume par unité de temps et de surface) permettent d'augmenter la production de bois, mais aussi de raccourcir la révolution et par conséquent de réduire le risque d'exposition aux aléas climatiques. Les gains en rectitude permettent l'amélioration du rendement au sciage du bois, et du rendement papetier grâce à la diminution du bois de compression<sup>6</sup>.

Depuis quelques années, d'autres caractères d'intérêt sont étudiés et pourront être intégrés prochainement comme critères de sélection dans le programme d'amélioration du pin maritime, comme par exemple :

- ✓ La tolérance à la sécheresse, dans le cadre du changement climatique.
- ✓ La résistance aux bioagresseurs (fomès, armillaire, rouille courbeuse, nématode, etc.)

---

<sup>4</sup> FCBA : Institut technologique Forêt Cellulose Bois-construction Ameublement

<sup>5</sup> CTBA : Centre technique du Bois et de l'Ameublement

<sup>6</sup> Un arbre courbé génère, dans la partie comprimée du tronc, du bois de compression destiné à réorienter l'arbre dans sa position verticale ; ce bois présente des cernes anormalement larges, des densités anormalement élevées, une accumulation de lignine et se rétracte fortement dans la direction longitudinale lors du séchage (GIS PMF 2002).





- ✓ La qualité du bois, et notamment la densité du bois. En effet, elle est corrélée avec les qualités technologiques traditionnelles du bois employé à l'état massif (sciages, déroulage) ou même à l'état reconstitué (panneaux), ainsi qu'avec le rendement en pâte à papier et avec le rendement énergétique. D'autres critères pourront aussi être pris en compte, comme par exemple l'angle du fil (orientation des éléments axiaux du bois ; le fil du bois est droit si les éléments axiaux sont disposés parallèlement à l'axe du tronc) ou la qualité de la branchaison (finesse et insertion à angle droit pour réduire la dimension des nœuds).

➤ **Contraintes spécifiques de l'amélioration des arbres forestiers**

Le programme d'amélioration du pin maritime est confronté aux contraintes spécifiques de l'amélioration génétique des arbres forestiers (Pilate *et al.* 2002) :

- Les superficies nécessaires à l'établissement des dispositifs, des vergers à graines et/ou des conservatoires sont très importantes (plusieurs centaines d'hectares).
- Les principaux caractères d'intérêt économique ne peuvent être évalués que tardivement (phénotypage à 8-12 ans pour les caractères de croissance chez le pin maritime). En effet, les corrélations juvénile-adulte pour ces caractères sont relativement faibles avant cet âge (Codesido *et al.* 2012 ; Kremer 1992 ; Lambeth 1980 ; Zas *et al.* 2004).
- L'intervalle entre les générations est très long, lié à la floraison tardive des arbres forestiers, même si des méthodes favorisant la floraison précoce ont été mises au point chez certaines espèces (Bramlett et Burris 1995 ; Bramlett 1997 ; Ross et Bower 1989). Chez le pin maritime, il faut attendre 7 à 8 ans (âge de la maturité sexuelle) pour réaliser les premiers croisements, et 10 ans pour obtenir les premières graines.
- Les caractères à améliorer sont très souvent polygéniques, fortement dépendants de l'environnement et ont une héritabilité faible à moyenne (0,1 à 0,6) (Burdon et Kumar 2004 ; Cornelius 1994 ; Falconer et Mackay 1996 ; Pâques 2013).
- L'aptitude à la propagation végétative (bouturage) du pin maritime décroît avec l'âge des arbres, ce qui limite la diffusion des meilleurs génotypes à un âge mature.

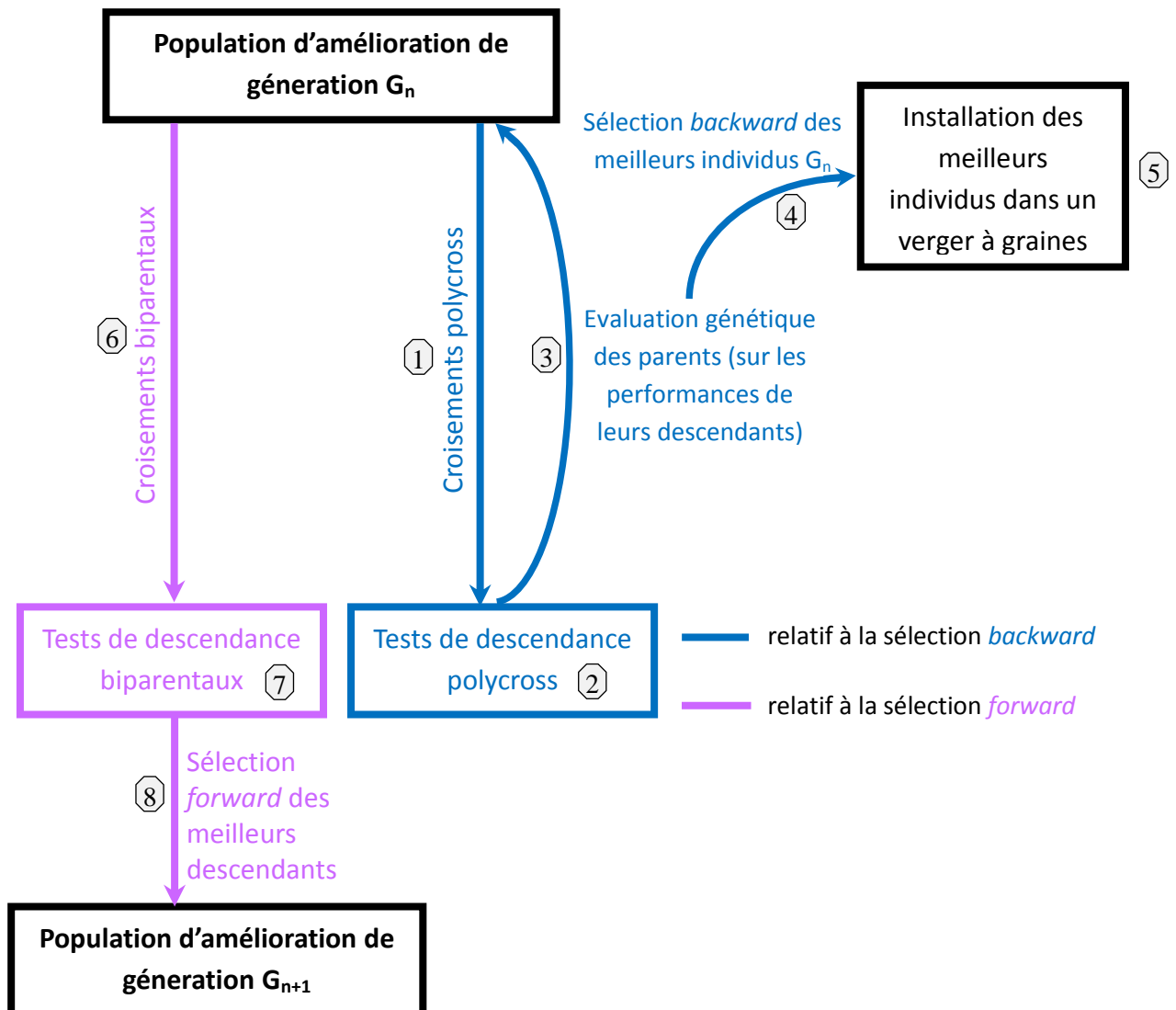
### 1.2.2 Une stratégie de sélection récurrente

L'amélioration du pin maritime repose sur une stratégie de sélection récurrente, comme c'est le cas dans les principaux programmes d'amélioration de conifères. Cette stratégie consiste en une succession de cycles associant une phase de sélection (pour sélectionner les individus les plus performants) et une phase de croisement (pour recombinaison des caractères favorables) (Namkoong *et al.* 1988 ; White 1987 ; Zobel et Talbert 1984). Chaque cycle implique la mise en place de parcelles d'essai sur le terrain pour tester les individus candidats à la sélection, le phénotypage et l'évaluation génétique de ces individus, la sélection des meilleurs génotypes ainsi que leurs croisements pour obtenir une descendance améliorée pour un ou plusieurs caractères d'intérêt.

**Figure 12 : Sélections backward et forward dans la stratégie de sélection récurrente du pin maritime**

D'une part, la stratégie de sélection backward : les valeurs génétiques des individus de la génération  $n$  ( $G_n$ ) sont estimées sur leur descendance. Après croisements polycross (1), des tests de descendance sont installés (2). Puis les descendants sont phénotypés, ce qui permet d'évaluer leurs parents  $G_n$  (3). Enfin les individus  $G_n$  ayant les meilleurs valeurs génétiques sont sélectionnés (4) et installés dans un verger à graines (5), qui produira la variété améliorée.

D'autre part, la stratégie de sélection forward. Les individus de la génération  $G_n$  sont croisés via des croisements biparentaux (6). Les meilleurs descendants sont sélectionnés (8) au sein des tests de descendance biparentaux (7) pour constituer la population d'amélioration de la génération suivante  $G_{n+1}$ .



Dans ce schéma de sélection, on distingue trois types de population :

- La population d'amélioration : il s'agit d'une population où les individus sont croisés et évalués. D'une part, des croisements (souvent de type polycross : croisements contrôlés entre une mère identifiée à évaluer et un mélange de pollens de différents pères appelé polymix) sont réalisés pour sélectionner les meilleures mères, évaluées sur les performances de leurs descendants (on parle alors de sélection *backward*, cf. Figure 12). D'autre part, les meilleurs parents (plusieurs centaines) sont croisés (croisements biparentaux) entre eux. Les meilleurs descendants sont sélectionnés (sélection *forward*) et constituent la population d'amélioration de la génération suivante (Figure 12).
- La population de production de la variété améliorée : les meilleurs génotypes de la population d'amélioration, issus de la sélection *backward*, sont sélectionnés pour l'installation de vergers à graines, qui produiront la variété améliorée commercialisée sous forme de jeunes plants.

Il existe deux grands types de vergers à graines pour le pin maritime :

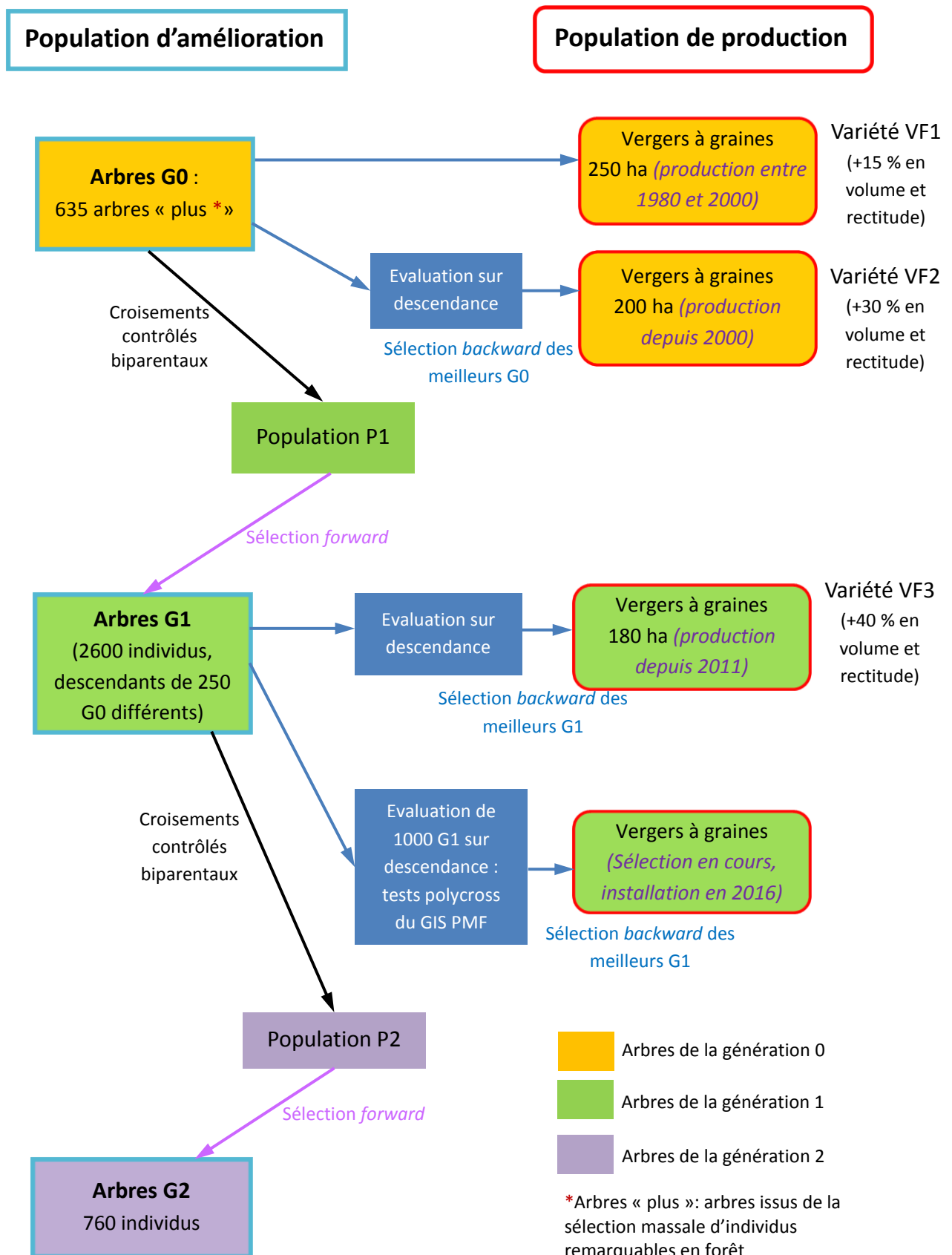
- Les vergers de clones : les composants sont des clones des individus sélectionnés (une quarantaine), multipliés par greffage de rameaux prélevés sur les arbres sélectionnés. La plupart des vergers installés en amélioration forestière sont de ce type.
  - Les vergers de semis (ou vergers de famille) : les composants sont des semis de descendance issues de croisements contrôlés entre les parents sélectionnés. Généralement, des croisements polycross sont réalisés : sur les fleurs femelles de chaque individu sélectionné, on applique un mélange de pollen contenant une grande majorité des pollens des individus sélectionnés.  
Ce type de verger est plus simple à installer (qu'un verger de clones) car il n'exige pas de greffes lentes et coûteuses. Cependant, sa fructification est plus tardive (8-10 ans contre 4-6 ans pour le verger de clones), et seule l'identité des mères des arbres constituant le verger est connue.
- La population de conservation : Une grande majorité des individus d'intérêt sont conservés par greffage au sein de parcs à clones. Chaque clone est représenté par plusieurs ramets<sup>7</sup>, identifiés par leur localisation (ligne-abscisse) dans le parc à clones.

---

<sup>7</sup> Ramets : copies clonales obtenues par voie végétative (généralement par greffage) à partir d'un seul individu originel (appelé ortet).

**Figure 13 : Historique simplifié du programme d'amélioration du pin maritime en Aquitaine**

La population  $P_n$  est constituée des arbres issus de croisements contrôlés entre individus de la génération  $n-1$ . Dans cette population  $P_n$  (tests de descendance biparentaux), les arbres sélectionnés sont nommés arbres  $G_n$ . Les gains des variétés améliorées (appelées VF pour Vigueur-Forme) sont exprimés en pourcentage par rapport à du matériel témoin non amélioré.



### 1.2.3 Constitution des populations d'amélioration successives

Les premiers travaux relatifs à la génétique du pin maritime en Aquitaine remontent à la fin des années 1930, avec l'installation d'essais de comparaison de provenances de l'aire naturelle du pin maritime. Ces essais ont mis en évidence la supériorité de la provenance landaise tant pour sa croissance rapide que pour sa bonne adaptabilité vis-à-vis des conditions écologiques contrastées présentes dans le massif des Landes de Gascogne (sols sableux podzolisés pauvres et acides, gorgés d'eau en hiver, froids hivernaux importants, sécheresse estivale) (Illy 1966). C'est pourquoi la provenance locale de la forêt d'Aquitaine a été retenue pour initier le programme d'amélioration du pin maritime dans les années 60, même si cette provenance présentait des défauts majeurs de forme (sinuosité basale du tronc).

Dans les années 1960-1970, la population de base du programme d'amélioration génétique a été constituée par la sélection phénotypique (sélection massale) d'arbres remarquables lors de deux campagnes de sélection successives au sein de tous les peuplements équiens de plus de 30 ans de la forêt landaise. Au total, 635 « arbres plus » présentant des caractéristiques de vigueur et de forme supérieures à la moyenne de leurs proches voisins, ont été sélectionnés. Ces arbres, considérés comme non apparentés, sont les arbres fondateurs du programme d'amélioration : ils constituent la génération G0 (Figure 13).

Ces arbres « plus » ont ensuite été évalués par des tests de descendance (dispositifs expérimentaux regroupant sur un même site les descendants de différentes familles) afin d'estimer leur valeur génétique pour les principaux caractères d'intérêt : volume, rectitude, branchaison. Les meilleurs individus G0 ont été sélectionnés, multipliés par greffage ou croisés entre eux (via des croisements polycross) et installés dans des vergers à graines de clones ou de familles (pour la production de la variété améliorée) (Figure 13).

Par ailleurs, des individus G0 ont été croisés entre eux (croisements biparentaux) et les meilleurs descendants ont été sélectionnés pour constituer la nouvelle population d'amélioration G1 (environ 2600 arbres, descendants de 250 G0 différents). Le même travail a été ensuite réalisé sur la population G1 (croisements, sélection) dans le cadre d'un nouveau cycle de sélection.

D'une part, après la mise en place du GIS PMF, de nombreux tests de descendance polycross ont été installés sur près de 100 ha (entre 1994 et 2003) dans le but d'évaluer près de 1000 G1 de provenance landaise. Deux polymix (mélange d'une quarantaine de pollens connus) différents ont été utilisés lors des croisements, correspondant à deux stratégies différentes des deux organismes sélectionneurs avant la fusion des programmes. Seules 18 mères ont été croisées avec les 2 polymix. Chaque test de descendance polycross comporte 35 blocs complets à placettes unitaires mono-arbres. Chaque bloc comprend, en plus des familles à tester, des témoins non améliorés issus de peuplements classés (constants dans tous les tests) et des lots ponts (familles choisies au hasard installées dans tous les tests). Chaque famille est représentée par 35 individus (à raison d'un arbre par bloc), et leur répartition à l'intérieur des blocs est aléatoire. L'évaluation des géniteurs (mères G1) est effectuée simultanément sur trois stations différentes pour explorer les trois grands types de landes représentatifs du massif aquitain (lande sèche, semi-humide et humide) et évaluer l'interaction génotype-environnement. De plus, les répétitions sur plusieurs sites permettent de pallier en partie aux risques accidentels de destruction des dispositifs expérimentaux (incendie, maladies, tempête, *etc.*). C'est sur l'un de ces tests de descendance polycross que mon travail de recherche a été réalisé.

D'autre part, près de 400 G1 ont été croisés deux à deux via un plan de croisement « double paire » (chaque parent est représenté dans deux croisements différents) pour constituer la génération suivante après sélection (G2).

La population d'amélioration est donc aujourd'hui constituée de 3 générations successives (notées G0, G1, et G2).

#### 1.2.4 Evaluation génétique des individus de la population d'amélioration

L'amélioration génétique repose sur la variabilité génétique, qui est, avec l'effet des variations environnementales, à la base des variations phénotypiques des caractères d'intérêt. Ces variations phénotypiques sont soit :

- discrètes : le nombre de modalités différentes est fini et souvent faible, et les individus peuvent être classés en catégories distinctes. Le plus souvent, il existe une relation simple entre le phénotype et le génotype.
- continues : les individus d'une population se répartissent alors de façon continue dans une certaine gamme de valeurs. C'est le cas pour la majorité des variations observables chez des organismes vivants, comme par exemple le poids, la taille, le rendement, *etc.* L'interprétation génétique de la variation de ces caractères est alors plus compliquée que dans le cas précédent : on est souvent incapable de déceler les gènes en ségrégation responsables de la variation observée, et on ne peut associer sans ambiguïté un phénotype donné à un génotype particulier. Les différences quantitatives dépendent de plusieurs gènes à différents locus et les effets de chacun ne sont pas distinguables facilement.

##### ➤ **Modèle infinitésimal**

La majorité des caractères d'intérêt pour la sélection sont des caractères quantitatifs, contrôlés par un très grand nombre de gènes indépendants à faible effet individuel. L'évaluation génétique des individus repose sur le modèle infinitésimal de Fisher, à partir des performances propres des individus à évaluer et/ou des performances mesurées sur des individus apparentés. On considère que les valeurs phénotypiques d'une population suivent une loi normale. On peut décomposer, de façon théorique, le phénotype (P) d'un individu selon :

- l'effet de l'environnement (E)
- l'effet génétique, appelé valeur génotypique (notée G), dû à l'assemblage des gènes de l'individu.

Ainsi, on note :  $P = G + E$

De même, la variation de ces valeurs peut être exprimée par leur variance  $\sigma^2_P, \sigma^2_G, \sigma^2_E$ .

Si on suppose les variables G et E indépendantes (on néglige ici les effets d'interaction GxE), on a  $\sigma^2_P = \sigma^2_G + \sigma^2_E$ .

Par ailleurs, la valeur génotypique G se décompose elle-même en deux types d'effets génétiques:

- **Effets additifs**, ou effets spécifiques à chaque gène, qui se transmettent de génération en génération. Chez un individu, leur cumul pour un caractère s'appelle valeur génétique additive, notée A. En moyenne, lors de la méiose, chaque gamète transmet la moitié de la valeur génétique additive du géniteur.

Ainsi en moyenne, un descendant reçoit la moitié de la valeur génétique additive de chacun de ses parents. Cette valeur correspond aussi à l'écart entre l'espérance phénotypique de l'individu et l'espérance phénotypique de la population (Falconer et Mackay 1996).

- **Effets non additifs** ou d'interaction. C'est l'ensemble des effets d'interaction entre allèles présents à un même locus (dominance, notée D) et entre allèles de locus différents (épistasie, notée I).

On a :  $G = A + D + I$  avec A, D et I supposés indépendants, et  $\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$ .

Il s'agit ici du modèle polygénique de base, qui permet de concilier les lois de Mendel et les observations faites sur les caractères quantitatifs. Il n'a pas de prétention explicative au niveau des mécanismes moléculaires en jeu.

Dans le cas du pin maritime, le gain génétique est diffusé au travers de variétés améliorées synthétiques (les plants améliorés résultent de croisements au hasard entre les individus sélectionnés constituant le verger à graines). C'est pourquoi la sélection ne porte que sur les effets additifs. Les effets non additifs peuvent être valorisés dans d'autres contextes, dans le cas de variétés issues de croisements contrôlés biparentaux (par exemple, pour produire une descendance hybride) ou de variétés clonales (multiplication en masse d'un individu d'intérêt) comme par exemple les variétés de peupliers, de saules, de cryptomeria, ou d'eucalyptus. Chez le pin maritime, la multiplication végétative n'est pas utilisée pour les sorties variétales car elle est techniquement complexe et pas encore économiquement rentable.

### ➤ **Ressemblance entre apparentés et covariances phénotypiques**

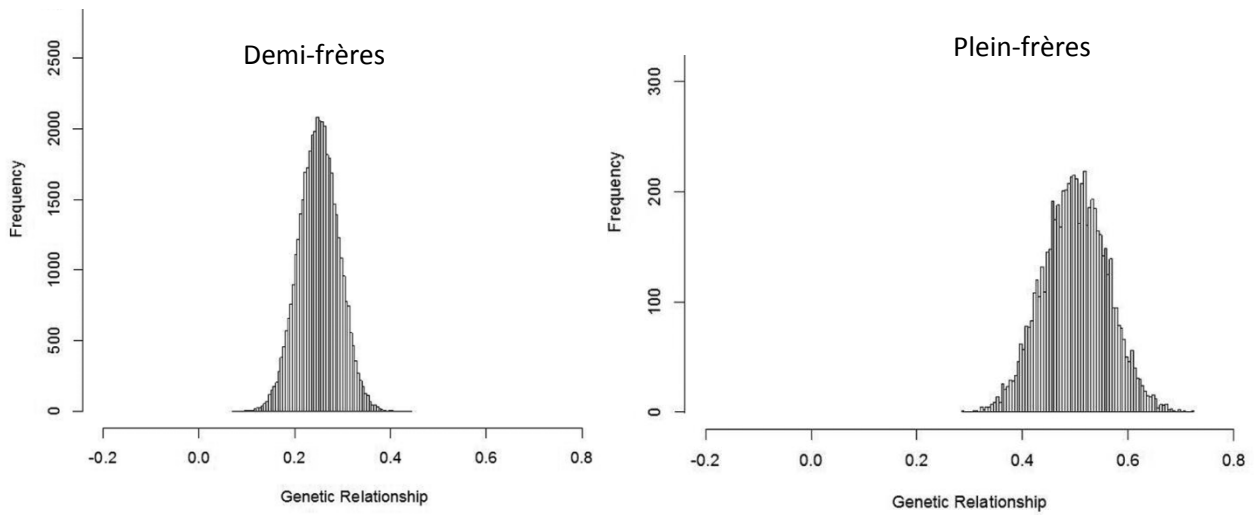
Le degré de ressemblance phénotypique entre deux individus, apprécié par une corrélation, dépend du caractère considéré, de leur lien de parenté et d'un éventuel environnement commun. Deux individus apparentés se ressemblent plus que deux individus choisis au hasard dans la population car ils possèdent statistiquement des gènes en communs et/ou un environnement commun. C'est sur ce partage de gènes que se base le calcul des covariances génétiques entre individus, permettant d'estimer leurs valeurs génétiques à partir de leurs valeurs phénotypiques.

Le calcul de l'apparentement fait intervenir les probabilités d'identité des allèles. Deux individus sont apparentés s'ils possèdent à un locus donné des allèles identiques provenant d'un ancêtre commun (appelés allèles IBD : *identity by descent*). Le coefficient de parenté  $f_{ij}$  entre deux individus  $i$  et  $j$  (Malécot 1948 ; Wright 1922) indique la probabilité d'obtenir deux allèles IBD lors du tirage aléatoire d'un allèle à un même locus chez  $i$  et  $j$ . Ce coefficient mesure la proportion attendue d'allèles IBD entre  $i$  et  $j$ . Par exemple,  $f_{ij}$  est égal à 0 entre deux individus non apparentés (*i.e.* n'ayant aucun ancêtre commun), à 0,125 entre deux individus ayant un seul parent commun (*i.e.*  $i$  et  $j$  sont demi-frères), à 0,25 entre deux pleins-frères (issus de parents non apparentés) ou entre un parent et son descendant.

Cependant, il faut garder à l'esprit que le calcul des coefficients de parenté repose sur la connaissance du pedigree et la probabilité d'avoir des allèles IBD. Par conséquent, la parenté théorique attendue peut différer de la parenté réelle entre deux individus. En effet, la proportion effective d'allèles IBD entre deux individus dépend de la ségrégation mendélienne lors de la méiose.



Figure 14 (issue de Munoz et al. 2014) : Distribution des coefficients d'apparements génomiques dans une population de demi-frères et plein-frères, estimés à l'aide de 2182 marqueurs moléculaires SNPs, autour de la moyenne attendue (0,25 pour les demi-frères, 0,5 pour les plein-frères).



Par exemple, si l'on considère un seul locus, la parenté réalisée entre deux pleins-frères issus de parents non apparentés sera de 0 dans 25 % des cas, 0,25 (un allèle IBD) dans 50 % des cas, et 0,5 (deux allèles IBD) dans 25 % des cas. En moyenne la parenté est bien de 0,25. Cette variation diminue lorsqu'on augmente le nombre de locus considérés. Des études de génomique estiment ce degré d'apparentement à l'aide d'un très grand nombre de marqueurs moléculaires jalonnant le génome : elles montrent que la proportion d'allèles partagés entre individus suit une distribution normale autour de la moyenne attendue (Figure 14).

Les évaluations génétiques se basent sur cet apparentement statistique. La covariance entre les phénotypes des individus  $i$  ( $P_i$ ) et  $j$  ( $P_j$ ) est directement fonction de leur degré d'apparentement :

$$\text{Cov}(P_i, P_j) = A(i,j) \times \text{Var}(G),$$

où  $\text{Var}(G)$  est la variance génétique et  $A(i,j)$  est le coefficient d'apparentement entre  $i$  et  $j$ , égal au double du coefficient de parenté. Ce coefficient est utilisé pour produire la matrice d'apparentement additif  $\mathbf{A}$  (*additive relationship matrix* ou *numerator relationship matrix*). C'est une matrice symétrique carrée dont la diagonale contient les coefficients d'auto-apparentement ( $2f_{ii}$ ) et les éléments hors-diagonale correspondent aux coefficients d'apparentement ( $2f_{ij}$ ).

Enfin, tout comme l'estimation des paramètres génétiques, les coefficients de parenté et de consanguinité font implicitement référence à une population de base composée d'individus fondateurs, supposés non apparentés entre eux et non consanguins.

#### ➤ **Evaluation génétique et analyse de variance**

Jusqu'au début des années 2000, les analyses de génétique quantitative forestière reposaient sur l'estimation de la covariance entre individus d'une même famille et la décomposition de la variance observée en différentes composantes. Ainsi, les paramètres génétiques et les valeurs génétiques étaient estimés à l'aide de tests de descendance issus de croisements polycross ou de croisements biparentaux. Ces tests sont généralement constitués d'une centaine de familles de natures différentes (demi-frères et/ou plein frères) selon le plan de croisement considéré (polycross, biparentaux hiérarchiques, biparentaux factoriels), et d'une cinquantaine d'individus par famille. Les individus sont répartis dans des blocs pour prendre en compte la variabilité de l'environnement.

Les variances sont estimées à partir des espérances mathématiques des valeurs des carrés moyens selon un modèle linéaire adapté. Par exemple, dans un test de descendance polycross composé de familles de demi-frères, les individus issus d'une même famille (*i.e.* même mère) ont en espérance 25 % de gènes identiques (IBD). On montre que la covariance phénotypique entre deux demi-frères  $i$  et  $j$  pris au hasard est égale au quart de la variance additive de la population, soit :

$$\text{Cov}(P_i, P_j) = \frac{1}{4} \sigma_A^2, \text{ avec } \sigma_A^2 \text{ la variance génétique additive.}$$

De plus, par analyse de variance, on peut décomposer la variance phénotypique totale  $\sigma_P^2$  en une variance inter-famille ( $\sigma_{Fam}^2$ ) et une variance intra-famille (variance résiduelle  $\sigma_e^2$ ). On montre que :

$$\text{Cov}(P_i, P_j) = \sigma_{Fam}^2.$$

On peut donc calculer la part du génotype transmissible à la descendance dans la variation totale du caractère d'intérêt. C'est l'héritabilité génétique (sens strict), définie par:

$$h_A^2 = \frac{\sigma_A^2}{\sigma_P^2} = \frac{4 \sigma_{Fam}^2}{\sigma_{Fam}^2 + \sigma_e^2}$$

L'héritabilité correspond aussi à la pente de la régression entre le phénotype moyen des parents et le phénotype moyen de leurs descendants. Il ne s'agit pas d'une valeur absolue : elle dépend de la population d'étude et de l'environnement au sens large.

Une fois les paramètres génétiques estimés, on peut prédire la valeur génétique des individus dans l'objectif de les classer et de réaliser une sélection. Soit  $\hat{A}_i$  l'estimation de la valeur génétique de l'individu  $i$  (appelée aussi Best Linear Predictor, BLP). Soit  $P_i$  la moyenne des phénotypes des  $n$  descendants de l'individu  $i$ . Soit  $\mu$  la moyenne phénotypique de la population.

$$\hat{A}_i = 2 \left( \frac{n}{n+\lambda} \right) x (P_i - \mu) \text{ avec } \lambda = \frac{4}{h^2} - 1$$

Ces méthodes génétiques sont très efficaces et ont été largement utilisées pour estimer les paramètres génétiques et prédire les valeurs génétiques à l'aide de nombreux tests de descendance. Toutefois, elles présentent certaines limites énumérées ci-dessous (Bidanel *et al.* 1990) :

- l'estimation des paramètres génétiques ne permet de prendre en compte qu'un seul type d'apparement par test analysé,
- les effets environnementaux sont difficiles à prendre en compte,
- les dispositifs ne doivent pas être trop déséquilibrés (*i.e.* avec des effectifs inégaux par modalités : famille, conditions de milieu, phénotypage),
- le mode de correction des effets du milieu rend presque impossible la comparaison d'indices de sélection d'individus issus de tests différents,
- il n'est donc pas possible de combiner des informations d'individus apparentés mais appartenant à des générations ou à des tests différents,
- les indices de sélection ne permettent pas de prendre en compte l'évolution de la variance génétique des caractères au cours des générations due à la sélection et à d'éventuels croisements préférentiels.

Dans le programme d'amélioration du pin maritime, la sélection des individus G1 s'est basée sur cette analyse de variance réalisée à l'aide du logiciel Diogene (Baradat 2000). Tout d'abord, on résout les équations d'un modèle à effets fixes pour estimer les effets blocs et familles (voire l'interaction bloc-famille) et on ajuste les performances des individus en fonction des effets blocs. Puis on résout les équations d'un modèle à effets aléatoires pour estimer les valeurs génétiques maternelles (et parfois paternelles dans le cas de tests de descendance issus de croisements factoriels). Enfin les meilleurs G1 sont sélectionnés sur un index multicritères incluant leurs valeurs génétiques pour la croissance et la rectitude.

### ➤ **Evaluation génétique et BLUP : le modèle mixte de Henderson**

Le développement récent des outils informatiques a permis l'utilisation de nouveaux modèles permettant de considérer dans une même analyse les effets environnementaux et les effets génétiques. L'introduction du modèle mixte par Henderson (Henderson 1950, 1975) permet l'estimation de ces effets par l'approche du maximum de vraisemblance.

Cette méthode a été d'abord très utilisée en génétique animale sous la forme du modèle individuel (appelé aussi modèle animal), dès les années 80. Son utilisation en génétique forestière est beaucoup plus récente, mais devient indispensable au vu de l'avancement des programmes de sélection (multi-génération, multi-sites, plusieurs types de tests) et de la bonne connaissance du pedigree des arbres des populations d'amélioration.

Dans les dispositifs des évaluations génétiques, le modèle mixte est utilisé pour estimer les effets environnementaux (effets fixes) et prédire les valeurs génétiques (effets aléatoires) à partir d'un vecteur de données (contenant les performances des individus).

Sous forme matricielle, le modèle linéaire mixte gaussien se définit de la manière suivante :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \text{ avec}$$

- $\mathbf{y}$  ( $n \times 1$ ) le vecteur des  $n$  observations,
- $\boldsymbol{\beta}$  ( $p \times 1$ ) le vecteur des  $p$  effets fixes et  $\mathbf{X}$  ( $n \times p$ ) sa matrice d'incidence,
- $\mathbf{u}$  ( $q \times 1$ ) le vecteur des  $q$  effets aléatoires et  $\mathbf{Z}$  ( $n \times q$ ) sa matrice d'incidence,
- $\mathbf{e}$  ( $n \times 1$ ) le vecteur des erreurs résiduelles, avec  $n$  le nombre d'observations, et  $p$  et  $q$  le nombre d'effets fixes et aléatoires à estimer, respectivement.

Ce modèle suppose que  $u$  et  $e$  suivent des lois normales indépendantes. Les espérances de  $u$  et  $e$  sont supposées connues ( $E[u] = E[e] = 0$ ) ainsi que les matrices de variance-covariance des effets aléatoires :

- $\text{Var}(\mathbf{u}) = \mathbf{A}\sigma_u^2 = \mathbf{G}$ , où  $\mathbf{A}$  est la matrice de parenté de la population basée sur la connaissance du pedigree (dont l'élément  $(k,l)$  est égal au coefficient de parenté entre les individus  $k$  et  $l$ )
- $\text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2 = \mathbf{R}$  (avec  $\mathbf{I}$  matrice identité)
- et  $\text{Cov}(\mathbf{u}, \mathbf{e}) = 0$

On a alors  $\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$  et par construction  $E[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$  où  $\boldsymbol{\beta}$  est un vecteur d'inconnues.

Les composantes de la variance doivent être estimées avant de pouvoir obtenir les solutions pour les effets fixes  $\hat{\boldsymbol{\beta}}$  (BLUE, *Best Linear Unbiased Estimators*) et aléatoires  $\hat{\mathbf{u}}$  (BLUP, *Best Linear Unbiased Predictors*). La méthode du maximum de vraisemblance restreint (REML) estime les variances qui renvoient à la population de base (dont descendent les individus observés). Henderson (1963) a montré que les estimations des effets fixes et les prédictions des valeurs génétiques peuvent s'obtenir comme solutions du système suivant :

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_u^2}\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

$$\text{Avec } \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \text{ et } \hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

#### ➤ Propriétés de l'évaluation génétique à l'aide du modèle mixte (« BLUP »)

La méthode « BLUP » présente de nombreux avantages comparée à l'analyse de variance « classique » :

- elle utilise la covariance additive entre tous les individus observés, quel que soit leur niveau d'apparentement, ce qui permet d'augmenter le nombre de données par individus et donc d'estimer plus précisément les valeurs génétiques
- elle permet une estimation simultanée des effets fixes et aléatoires : ainsi l'ajustement au milieu prend en compte la valeur génétique estimée dans chaque milieu
- elle gère facilement les dispositifs expérimentaux déséquilibrés

- elle est flexible et peut tenir compte de nombreux effets, comme par exemple les corrélations génétiques et résiduelles entre caractères, les mesures répétées (sur plusieurs années), les groupes génétiques, les variances hétérogènes, les interactions entre génotype et environnement, *etc.*
- elle permet de combiner des informations d'individus apparentés appartenant à des dispositifs différents (analyses inter-sites)

Ainsi, cette méthode permet d'évaluer la valeur génétique d'un individu en utilisant les performances de tous ses apparentés (quel que soit leur dispositif d'évaluation, leur génération), mais aussi de comparer des valeurs génétiques entre individus, entre tests, entre générations.

Au sein du programme d'amélioration du GIS PMF, les valeurs génétiques des individus sont estimées par la méthode du BLUP depuis 2011. Des données phénotypiques ont été accumulées depuis les années 60s dans le cadre du programme d'amélioration du pin maritime dont la population est désormais constituée de trois générations (G0, G1 et G2). Il existe de nombreuses connexions entre ces données phénotypiques via le pedigree des individus et les corrélations entre caractères. Une évaluation BLUP de la population d'amélioration prend donc tout son sens dans ce contexte. Aujourd'hui, les évaluations génétiques se font via un modèle mixte individuel. Ce nouveau système d'évaluation a été mis en place par l'INRA avec le STBA (Southern Tree Breeding Association) en Australie. Il se compose d'une base de données (Dataplan) et d'un calculateur BLUP (Treeplan, développé par PlantPlan (McRae *et al.* 2004)). Le couplage de la base de données et du calculateur permet l'actualisation des valeurs génétiques dès l'intégration de nouvelles mesures phénotypiques dans la base de données. Cette évaluation génétique est réalisée en trois grandes étapes :

- La première étape consiste à intégrer dans la base de données les données de chaque dispositif en renseignant notamment les caractéristiques du dispositif (localisation, design expérimental, matériel évalué, caractères mesurés...), le pedigree des individus et les valeurs phénotypiques individuelles (après une étape de validation des données brutes).
- La seconde étape est une analyse de chaque dispositif. Les données phénotypiques sont d'abord ajustées spatialement (modèle d'autocorrélation). Un modèle d'analyse multivarié permet ensuite de calculer les corrélations génétiques entre caractères. Enfin, les variances des caractères sont obtenues en fixant les corrélations génétiques (on utilise ici les corrélations génétiques de la population d'amélioration estimées à partir de l'ensemble des corrélations intra-dispositif).
- La troisième étape est l'évaluation BLUP de tous les individus de la population d'amélioration (reliés par un pedigree) à partir des mesures ajustées spatialement et standardisées par la variance génétique estimée dans chaque dispositif.

Les principaux dispositifs du programme d'amélioration ont été intégrés dans ce système d'évaluation, ce qui correspond à environ 70 dispositifs et plus de 450 000 arbres. Pour chacun de ces arbres, on dispose donc d'une valeur génétique pour l'ensemble des caractères mesurés : croissance (circonférence, hauteur), écart à la verticalité et qualité du bois (branchaison, densité du bois, angle du fil).

Une précision est associée à chacune de ces valeurs génétiques ; elle traduit la quantité d'information relative à l'individu évalué et à ses apparentés, que ce soit pour le caractère considéré ou pour les caractères corrélés.

### 1.2.5 Déploiement du gain génétique des variétés améliorées

Le gain génétique mesure la supériorité du matériel issu de sélection par rapport au matériel témoin non amélioré (GIS PMF 2002). Il est souvent exprimé en pourcentage par rapport au témoin. Avant l'installation d'un verger à graines (VG), le gain est estimé à partir des valeurs génétiques des individus constituant le verger. C'est le gain qui sera potentiellement réalisé par les descendants des individus sélectionnés placés dans le VG. Une fois que les performances de la variété améliorée (constituée par la descendance du VG) ont pu être mesurées concrètement sur le terrain et comparées aux descendants des témoins non améliorés (observations dans les mêmes conditions), le gain est dit réalisé. Cependant, le gain génétique est très variable selon l'environnement. De plus, la composition de la variété change selon les années de production (floraison et contributions des parents dans le VG, pollution pollinique très importante les premières années de production), même si son impact sur le gain est plus anecdotique. Ainsi, le gain génétique ne peut s'estimer qu'en moyenne et après avoir mesuré un très grand nombre de parcelles.

#### ➤ Compositions des vergers à graines

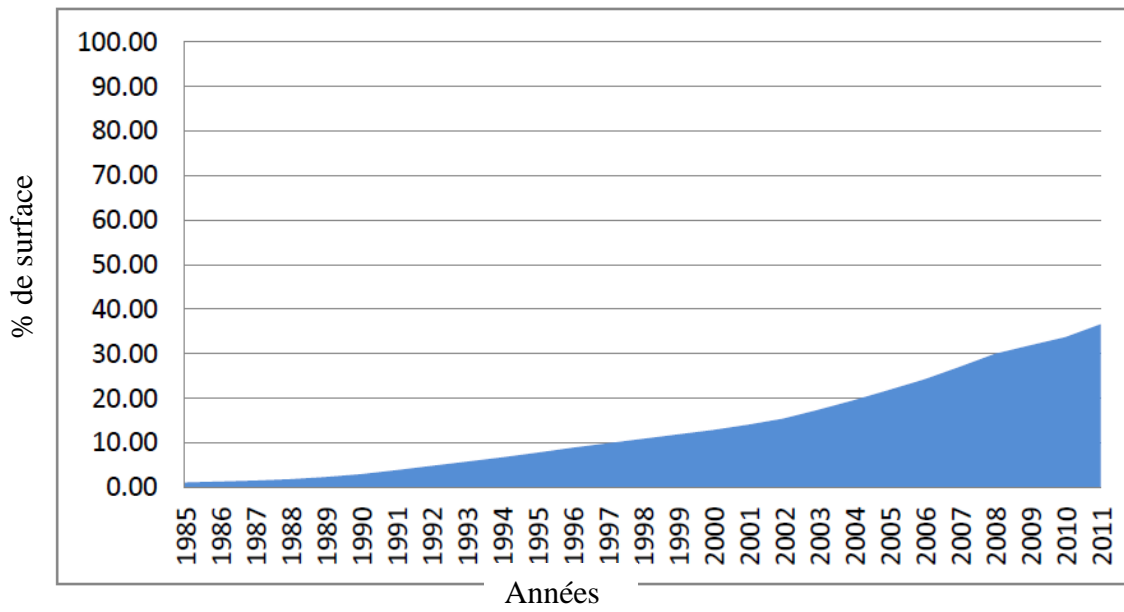
- Les premiers VG, dit de « première génération » : 250 ha de VG ont été installés entre 1962 et 1978. Deux VG sont des VG de clones installés au début des années 1960 et sont constitués d'un grand nombre d'arbres G0 greffés (arbres plus, issus de la sélection massale dans tout le massif landais). Puis entre 1967 et 1978, trois VG de familles ont été installés avec des semis provenant de croisements contrôlés entre près de 300 G0 sélectionnés (GIS PMF 2002).

Ces VG ont permis un gain génétique réalisé moyen de 15 % sur le volume (à 15 ans) et sur la rectitude. Ces gains permettent de raccourcir le cycle sylvicole de 4-5 ans.

- La « deuxième génération » de VG (200 ha installés entre 1986 et 1995) : ces VG bénéficient des connaissances acquises lors de l'évaluation sur tests de descendance des individus G0 et se caractérisent ainsi par une sélection plus intensive de leurs constituants. Trois VG de familles (de 60 ha chacun) ont été constitués (entre 1987 et 1991) avec des semis issus de croisements contrôlés entre 34 individus G0 ayant les meilleures aptitudes générales à la combinaison. De plus, à la fin des années 1980, trois VG de clones en pollinisation contrôlée ont été créés par greffage de 40-50 individus G0. Ici, contrairement au mode de reproduction libre utilisé habituellement en VG, l'inter-croisement entre les individus G0 au sein du VG est réalisé de façon contrôlée (GIS PMF 2002).

Ces VG ont donc une base génétique plus réduite (par rapport aux VG précédents), ce qui conduit à une meilleure homogénéité des parcelles reboisées et des qualités du bois produit. Cette sélection plus intensive permet d'espérer un gain génétique moyen de 30 % sur le volume et la rectitude (par comparaison au matériel non amélioré). Ces gains permettraient de raccourcir le cycle sylvicole de 6 à 8 ans. Ainsi la durée moyenne d'une rotation passe de 40 à 32 ans.

Figure 15 : Evolution du pourcentage de la surface du massif occupée par les variétés améliorées (d'après les EAB, Agreste) entre 1985 et 2011 (aire bleue)



- La « troisième génération » de VG (180 ha installés entre 2002 et 2010) est constituée de plus de 40 individus G1 sélectionnés (évalués dans des tests de descendance polycross). Ces VG (trois VG de clones et quatre VG de familles) produisent la variété landaise Vigueur-Forme de troisième génération (VF3). Le gain génétique moyen espéré est d'environ 40 % pour le volume et la rectitude (par comparaison au matériel non amélioré) (GIS PMF 2002).

➤ **Reboisement et besoins en graines améliorées**

La « première génération » de VG a permis de reboiser 80 000 ha par plantation entre 1980 et 2000. Depuis les années 2000, leur exploitation a cessé car les VG de 2<sup>ème</sup> génération (présentant des gains génétiques supérieurs) ont pris le relais (GIS PMF 2002). La commercialisation des variétés issues des VG de 2<sup>ème</sup> génération a débuté dans les années 1990. Les VG de 3<sup>ème</sup> génération sont entrés en production depuis 2011. A cause des forts besoins en reboisement post tempête de 2009 et des mauvaises fructifications de ces dernières années (faibles récoltes en graines), la production des VG de 3<sup>ème</sup> génération n'est pas suffisante et donc les VG de 2<sup>ème</sup> génération sont toujours en production.

Aujourd'hui, plus de 90 % des surfaces reboisées en pin maritime au sein du massif des Landes de Gascogne le sont avec des plants issus de graines améliorées récoltées en VG (100 % des plants sont issus de graines de VG). Actuellement, environ 40 % des surfaces du massif des Landes de Gascogne sont boisées avec des variétés améliorées (contre ~12 % en 2000) (Figure 15). Cette surface est en constante augmentation, au rythme du renouvellement du massif. Dans le contexte post-tempêtes actuel, les besoins en reboisement sont très importants et avoisinent actuellement les 20 000 ha/an, soit un besoin d'environ 40 millions de plants (équivalent à 3 tonnes de graines). Ce rythme de renouvellement offre la possibilité d'introduire plus rapidement du nouveau matériel végétal, plus performant et mieux adapté au contexte économique et à l'environnement actuel. Il constitue donc une opportunité pour mieux renouveler la forêt plantée du massif des Landes de Gascogne.

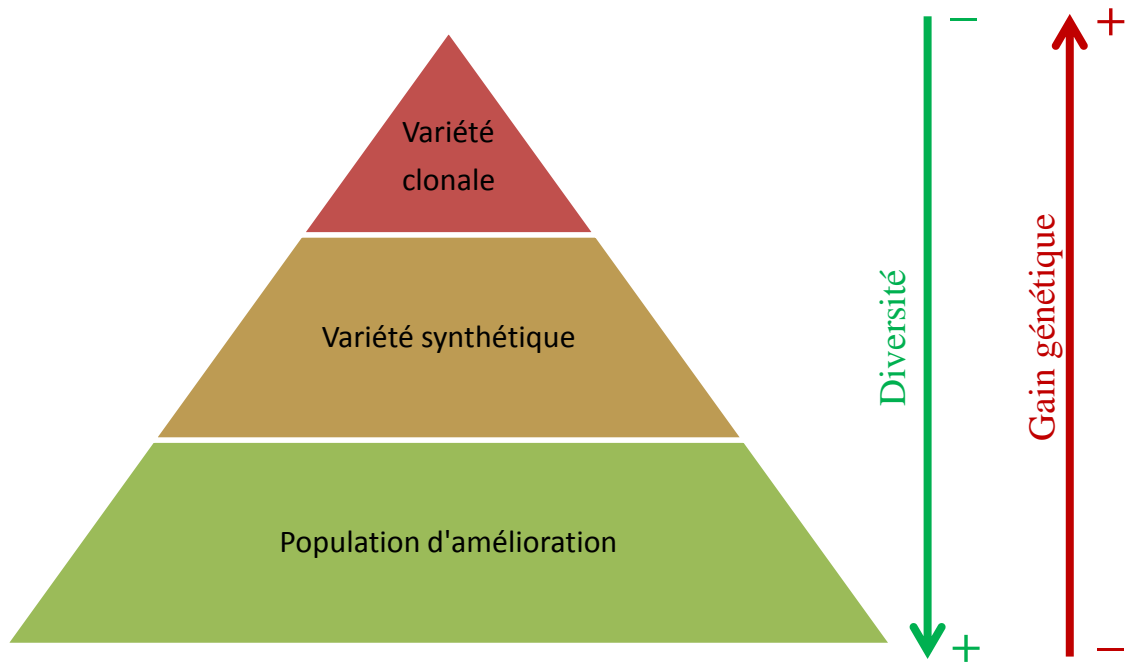
### 1.2.6 Gestion de la diversité génétique

La variabilité génétique au sein des populations d'arbres forestiers est, d'une façon générale, très importante (Hamrick *et al.* 1992). Cela s'explique par différents facteurs. Les arbres ont des populations de grande taille, ce qui favorise l'apparition des mutations (qui engendrent la diversité). Ils ont aussi de longues générations qui se mélangent, favorisant le maintien de ces mutations. Parmi les arbres forestiers, le pin maritime ne fait pas exception. De plus, son régime de reproduction préférentiellement allogame (favorisant les flux de gènes entre individus), sa dispersion pollinique à grande distance (jusqu'à 50 km (Illy et Sopena 1963)) et sa domestication très récente sont d'autres facteurs favorisant son importante diversité (GIS PMF 2014 ; Hamrick *et al.* 1992).

Cette variabilité génétique est un élément clé permettant d'assurer à la fois un progrès génétique sur le long terme, l'introduction de nouveaux critères de sélection, et de garantir les capacités d'adaptation face aux changements climatiques (Bouffier *et al.* 2009). Ainsi les sélectionneurs s'efforcent de concilier gain génétique et diversité à court et long terme.



*Figure 16 : Niveau de diversité et de gain génétique au sein des différentes populations d'un programme d'amélioration*



Pour cela, les populations sont subdivisées en strates de différents niveaux de diversité, depuis les populations de conservation (issues de différents peuplements provenant de l'aire de répartition de l'espèce) jusqu'aux variétés à moindre diversité et fort gain génétique (Cf. figure 16). Ces strates permettent de dissocier les objectifs de maintien de diversité sur le long terme et de création de gain génétique, et d'infuser du matériel génétique entre strates lorsque nécessaire (*i.e.* introduire dans une strate plus intensivement sélectionnée de nouveaux individus depuis une strate plus diversifiée mais de moindre gain génétique).

➤ **Gestion de la diversité au sein des populations de provenance landaise**

• A long terme, dans la population d'amélioration

Les populations d'amélioration doivent conserver une forte diversité génétique au fil des générations. En effet, la diversité est le moteur de l'amélioration et permet :

- d'assurer des progrès génétique sur le long terme,
- l'introduction de nouveaux critères de sélection dans les futures générations des populations d'amélioration, permettant l'adaptation à de nouvelles contraintes environnementales et économiques qui pourraient surgir (changement climatique, apparition de nouveaux pathogènes, *etc.*).

Pour maintenir des formes alléliques fréquentes à rares sur le long terme, la taille de population (par population d'amélioration et par zone de sélection) généralement préconisée est de l'ordre d'une centaine d'individus non apparentés (GIS PMF 2014). Dans le cadre du programme d'amélioration de la provenance landaise, 250 arbres fondateurs G0 (arbres « plus » représentant la variabilité naturelle de la provenance landaise) constituent la population d'amélioration de base. La taille de population d'amélioration landaise est supérieure aux standards de nombreux programmes de sélection (White *et al.* 2007).

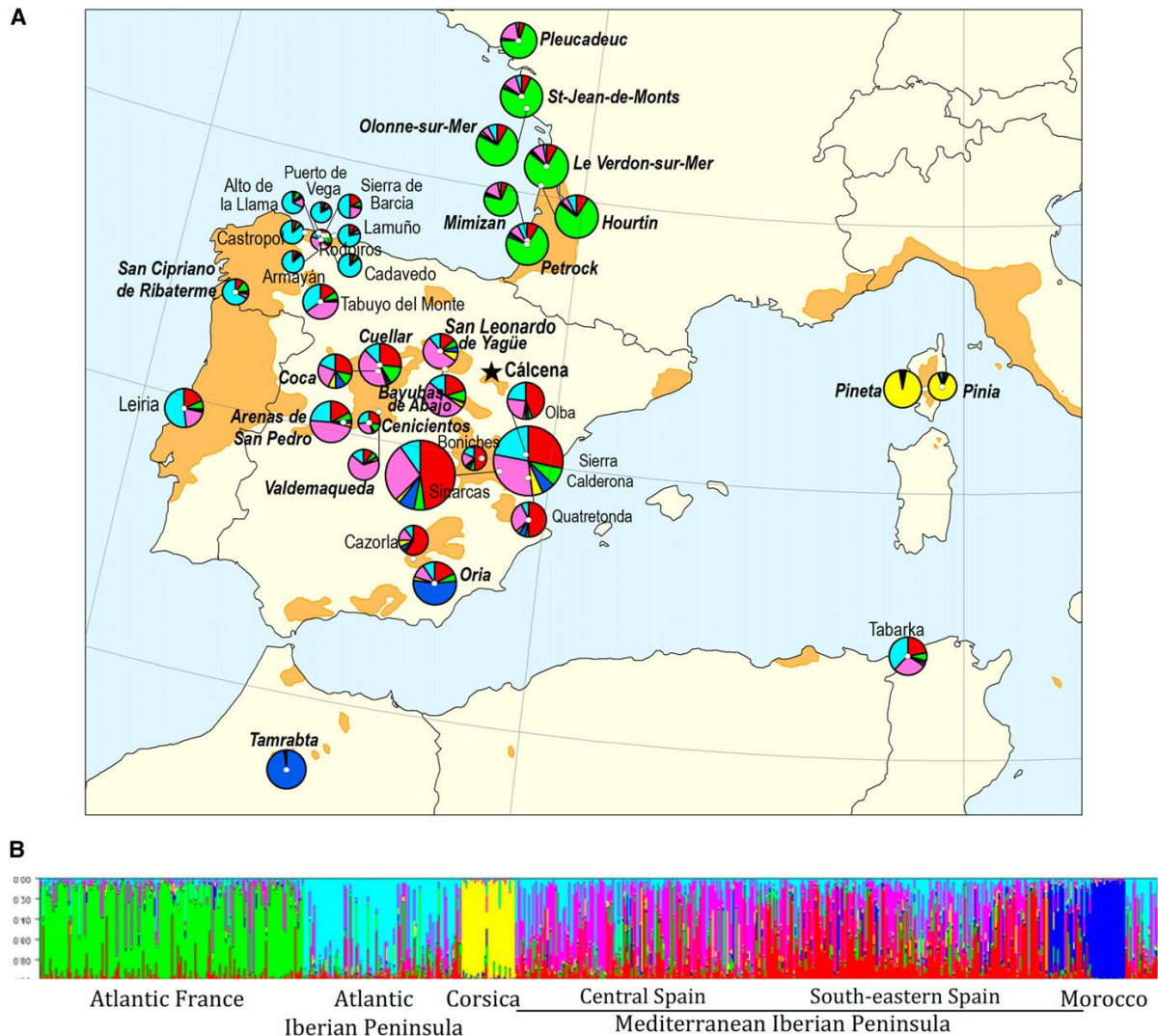
Concrètement, l'évolution de la diversité dans la population d'amélioration du pin maritime a été observée sur des caractères d'intérêt. Bouffier *et al.* (2008) ont montré que la variabilité génétique s'est maintenue au cours des générations de sélection pour un caractère non sélectionné (densité du bois), tandis que pour un caractère sélectionné (croissance), elle a d'abord diminué lors de la phase de recrutement de la population d'amélioration (sélection des G0 en forêt), puis elle s'est maintenue dans la génération suivante. On a donc pu vérifier que la diversité ne s'est pas érodée notablement dans les premières générations de sélection.

• A court terme, dans la population de production

Lors de la constitution de la population de production, l'enjeu est de produire une variété destinée aux sylviculteurs et aux industriels qui soit à la fois performante sur les caractères d'intérêt sylvicole (maximum de gain génétique), homogène et ayant un niveau suffisant de variabilité des critères adaptatifs pour supporter les variations environnementales classiques d'un cycle de révolution (Hosius *et al.* 2000 ; Lindgren *et al.* 2009 ; Stoehr *et al.* 2004). En effet, la révolution très longue (plusieurs décennies) des espèces forestières favorise leur exposition aux contraintes environnementales évolutives comme le changement climatique ou l'apparition de maladies et bioagresseurs (Hansen 2008 ; Johnson et Lipow 2002 ; Muller-Starck 1995). Cependant, parce qu'un peuplement issu de VG n'est installé que pour une seule génération (35 ans en moyenne pour le pin maritime), on peut accepter un niveau de diversité génétique moindre.

**Figure 17 (issue de Jaramillo-Correa et al. 2015) : Diversité génétique des différentes provenances géographiques du pin maritime**

(A) Distribution géographique des 6 groupes génétiques bien différenciés obtenus en géotypant 772 individus de 36 populations différentes avec 9 marqueurs SSRs nucléaires (microsatellites) (B). La zone orangée (A) représente l'aire de distribution naturelle du pin maritime.



Cette diversité minimum requise varie selon les espèces et dépend de la durée du cycle de révolution, de la variabilité du milieu à laquelle la variété devra faire face au cours de sa vie (Johnson et Lipow 2002) et du rythme de renouvellement des variétés (qui détermine la surface plantée avec une même variété au niveau du massif). En termes d'ordre de grandeur, Johnson et Lipow (2002) ont montré qu'une variété forestière composée d'une vingtaine de parents peu ou non apparentés serait suffisante pour égaler le niveau de risque face à des aléas d'un peuplement issu de graines non améliorées. Leur revue montre que les VG comportant 25 parents non apparentés retiennent 92 % de la variabilité génétique de la population naturelle d'origine.

Seuls les allèles de faible fréquence dans la population naturelle (allèles rares) seront probablement absents dans le VG. Cependant, ces allèles rares, même s'ils peuvent potentiellement permettre la résistance à un aléa, ont peu d'importance dans un peuplement car ils ne permettraient qu'à très peu d'arbres de survivre. De tels allèles rares sont par contre importants dans la population d'amélioration, où leur fréquence pourrait être augmentée par sélection et bénéficier aux futurs VG.

Selon les espèces et le type de variété produite, la diversité intra-variété peut être quasi nulle dans le cas d'une variété clonale, ou très importante dans le cas d'une variété synthétique provenant d'un VG composé d'un très grand nombre de clones, avant éclaircie génétique par exemple.

Pour le pin maritime, le gain génétique est diffusé au travers de variétés synthétiques, c'est-à-dire un mélange de familles dont les parents sont sélectionnés. Ainsi, chaque individu est une combinaison unique de gènes. Le niveau de diversité global obtenu dépend du nombre de parents dans le VG à l'origine de la variété, de leur participation homogène ou non à la production des semences, et du taux de pollution pollinique. De plus, il est nécessaire de limiter l'apparentement entre les parents sélectionnés pour l'établissement du VG afin de limiter les risques de dépression de consanguinité (Durel *et al.* 1996 ; Olsson *et al.* 2001 ; Stoehr *et al.* 2008). Actuellement, le nombre de parents en équivalents non-apparentés dans les VG est de l'ordre de 30, ce qui est un nombre important comparés aux standards internationaux pour des variétés forestières.

#### ➤ **Utilisation de différentes provenances pour une variabilité génétique plus large**

Des tests de provenances et des études sur la structure génétique des populations de pin maritime basées sur des marqueurs moléculaires (Bucci *et al.* 2007 ; Burban et Petit 2003 ; Jaramillo-Correa *et al.* 2015 ; Mariette *et al.* 2001 ; Santos-del-Blanco *et al.* 2012) révèlent une très grande variabilité génétique entre les différentes provenances du pin maritime. Cette variabilité s'explique notamment par son aire naturelle de distribution morcelée, dans des environnements très variés (Figure 17).

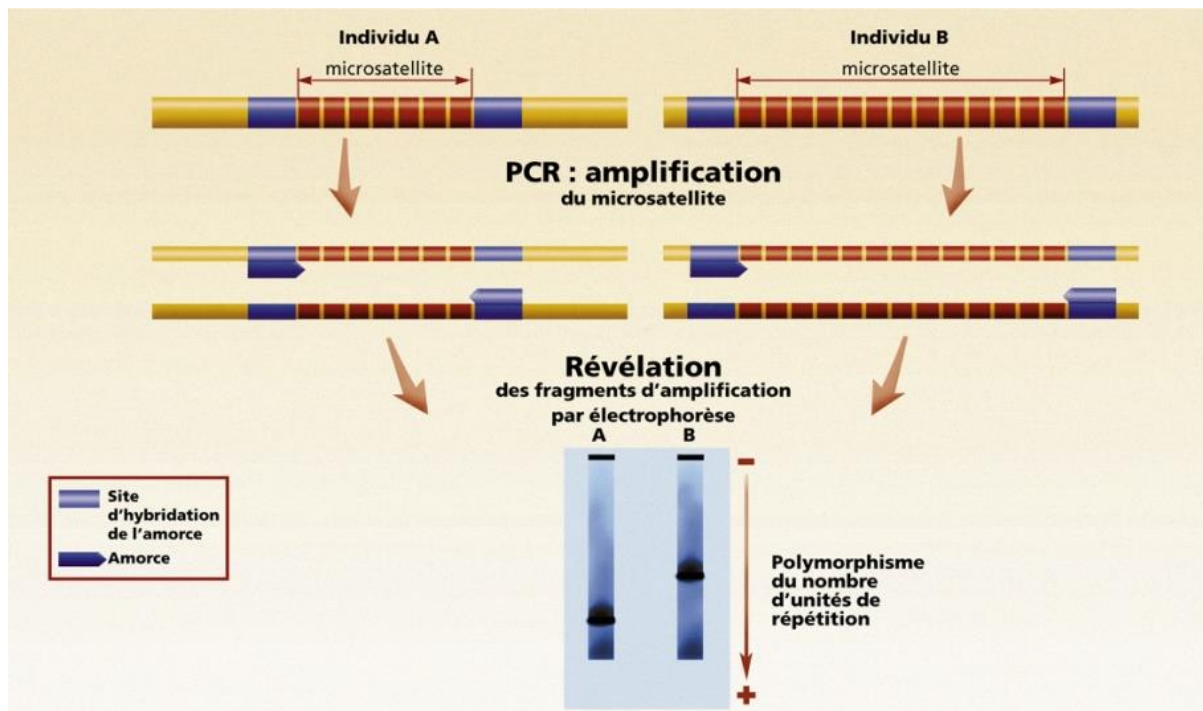
A côté de la provenance landaise, le programme d'amélioration du GIS PMF développe une population d'amélioration de provenance corse, particulièrement intéressante pour la qualité de la branchaison et la rectitude du tronc (GIS PMF 2014). Cette population est valorisée dans le cadre des variétés Landes x Corse. Une population de provenance marocaine a également été sélectionnée dans une perspective de sélection pour la résistance à la sécheresse et à certains pathogènes (*Matsucoccus* dans le sud-est, armillaire).



De plus, les provenances méditerranéennes de pin maritime, adaptées à des climats plus secs que le sud-ouest de la France, pourraient être intéressantes pour le reboisement dans le massif landais dans le cadre du changement climatique et de l'aggravation annoncée des sécheresses en Aquitaine. Cependant, des croisements (avec la provenance landaise locale résistante au froid) et un programme de sélection seront très probablement nécessaires pour garantir un niveau de résistance au froid suffisant. Un programme d'évaluation de croisements interprovenances (Corse, Portugal, Maroc) est d'ailleurs mené par le GIS GPMF (GIS PMF 2014).

**Figure 18 : Marqueurs microsatellites (SSR)**

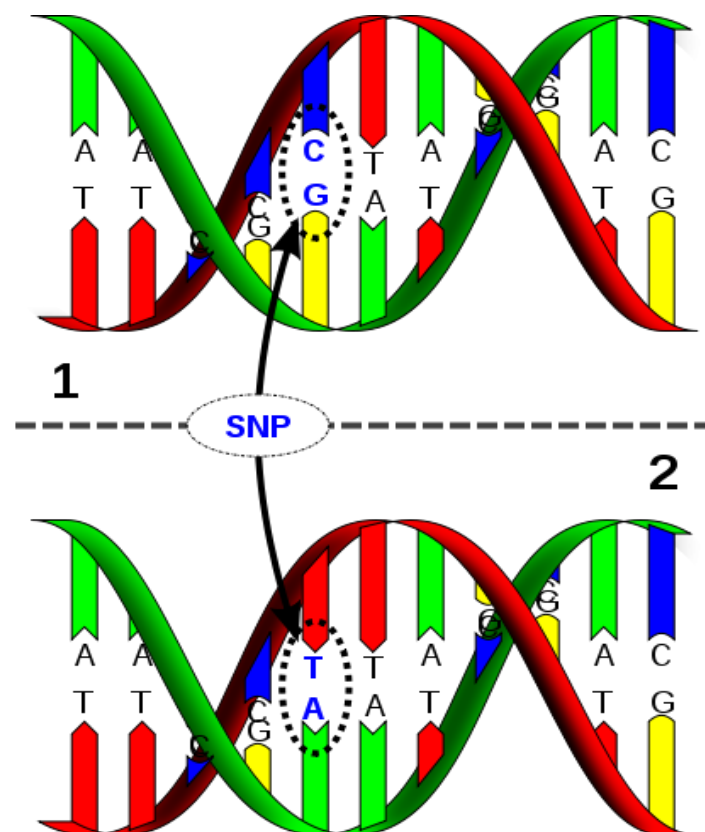
<http://www.gnis-pedagogie.org/biotechnologie-amelioration-marqueur-microsatellite.html>



**Figure 19 : Marqueurs SNP, variabilités ponctuelles de la séquence d'ADN.**

© David Hall / Licence Creative Commons

<http://www.futura-sciences.com/magazines/sante/infos/dico/d/genetique-snp-6348/>



## 1.3 Intégration des marqueurs moléculaires dans les programmes d'amélioration forestiers

### 1.3.1 Définitions et utilisations

Les marqueurs moléculaires sont des petits fragments d'ADN (codants ou non) présentant un polymorphisme selon les individus, ce qui permet de caractériser un génotype. La découverte des marqueurs moléculaires a permis aux biologistes d'étudier la transmission du génome au cours des générations, l'histoire évolutive des espèces, les flux de gènes dans l'espace et le temps, les phénomènes d'adaptation au sein de différentes populations, d'approfondir le fonctionnement du génome, *etc.* Le marqueur moléculaire « idéal » est héritable, polymorphe, co-dominant, reproductible, neutre, très fréquent tout au long du génome, et facilement déterminable. Il existe différents types de marqueurs moléculaires, mais les plus récents et les plus utilisés aujourd'hui sont les marqueurs microsatellites (aussi appelé SSR : *simple sequence repeat*) et les SNPs (*single-nucleotide polymorphism*). Le microsatellite est une séquence de di-, tri- ou tétra-nucléotides répétés en tandem (Figure 18). Ces éléments sont uniformément répartis en plusieurs copies sur l'ensemble du génome d'une espèce et présentent un taux de polymorphisme très élevé, qui repose sur la variation du nombre d'unités de répétition (la longueur de séquence). Les marqueurs microsatellites ont été longtemps utilisés pour reconstituer les pedigrees, car ils sont multialléliques, codominants et facilement observables (Jones *et al.* 2010). Le polymorphisme mononucléotidique (SNP) est la variation d'une seule paire de bases du génome (Figure 19). Malgré leur faible polymorphisme (généralement bi-allélique) les SNPs sont de plus en plus utilisés par la communauté scientifique car ils sont très abondants sur le génome, plus fiables, plus répétables, plus simples et moins coûteux à génotyper (Telfer *et al.* 2015), notamment grâce au développement des techniques de génotypage à haut débit (Mardis 2008) tel que Illumina golden gate assays, ABI TaqMan, Sequenom iPlex platforms (Ragoussis 2006).

Dans ce travail de thèse, ce sont les marqueurs SNPs qui ont été utilisés car on disposait d'une importante base de données, avec un grand nombre de marqueurs potentiels identifiés dans notre population d'amélioration (Chancerel *et al.* 2013) et d'une plateforme de génotypage sur place (<http://www.pgtb.u-bordeaux2.fr/>) dotée d'une technologie de génotypage des SNPs, le Sequenom® iPLEX® Gold (San Diego, CA, USA). Une étude basée sur des simulations suggère que 60 à 100 SNPs sont suffisant pour reconstituer avec précision le pedigree d'un millier d'individus (Anderson et Garza 2006). La capacité de discrimination entre les parents potentiels est fonction du nombre de marqueurs, de l'importance de la variabilité allélique à chaque locus, de la fréquence des allèles dans la population et du nombre de parents potentiels.

L'utilisation des marqueurs moléculaires est de plus en plus étudiée au sein des programmes d'amélioration génétique. Par exemple, de nombreux programmes de recherche sont menés pour tenter de sélectionner des individus sur un profil de marqueurs (*i.e.* leur génotype), notamment au travers de la sélection assistée par marqueurs ou de la sélection génomique. D'autres études portent sur l'utilisation de ces marqueurs comme outils d'aide à la gestion des populations d'amélioration (gestion de la diversité, structure des populations, reconstitution de pedigree, certification de variétés, *etc.*).



### 1.3.2 Des outils pour prédire le phénotype

#### ➤ Sélection Assistée par Marqueurs (SAM)

Le principe de la sélection assistée par marqueurs est de sélectionner des individus non plus à partir de données phénotypiques mais sur leur génotype aux marqueurs, en fonction des associations connues (basées sur le déséquilibre de liaison) entre les marqueurs et les QTL (Quantitative trait locus) d'intérêt. Les marqueurs permettent d'étudier la transmission d'un QTL, de sélectionner sur la présence d'un allèle d'un gène majeur (par exemple allèle conférant une résistance à une maladie), ou de présélectionner des individus porteurs des allèles recherchés.

L'information sur les marqueurs présente deux avantages majeurs. D'une part, elle peut être connue très tôt dans la vie des individus, ce qui permet de développer un critère de sélection précoce. D'autre part, elle permet de limiter les opérations de phénotypage. Ainsi la SAM peut se révéler très utile et permet notamment :

- de pallier un manque d'information phénotypique, par exemple, lorsque la mesure d'un caractère est très coûteuse, lorsqu'elle ne peut pas s'effectuer sur certaines catégories d'individus ou lorsqu'elle nécessite la destruction des individus à tester
- d'appliquer une plus grande intensité de sélection (plus d'individus génotypés que phénotypés)
- d'augmenter le gain par unité de temps.

La SAM est utilisée dans les programmes d'amélioration des grandes cultures, surtout pour le riz, le blé, le maïs et l'orge. Elle permet par exemple de suivre des gènes d'intérêt (gènes de résistance contre des agents biotiques, gènes liés à la productivité ou à la qualité), notamment lors de rétrocroisements (pour suivre l'introgression d'un gène) (Muranty *et al.* 2014). Elle est aussi très étudiée pour l'amélioration des bovins laitiers depuis le début des années 2000 (Fritz *et al.* 2003).

Chez les arbres forestiers, la découverte de QTL et leur utilisation en SAM ont fait l'objet de nombreuses études (comme par exemple Guimarães *et al.* 2007, Jermstad *et al.* 2001, Plomion *et al.* 1996b, Sewell *et al.* 2000, Wilcox *et al.* 2001). L'étude de l'architecture génétique des caractères quantitatifs permet de dénombrer et localiser les QTL sur le génome, et d'estimer leur effets (Chagne *et al.* 2003 ; Lepoittevin *et al.* 2012 ; Lerceteau *et al.* 2000 ; Marguerit *et al.* 2014 ; Neale et Savolainen 2004 ; Pot *et al.* 2006). La SAM semble intéressante chez les arbres forestiers qui ont des cycles d'amélioration très longs. Ces outils permettraient en effet de réduire ces cycles via une sélection précoce des individus sur la base de leur génotype.

Cependant, malgré de nombreuses études de détection de QTL, la SAM n'est pas utilisée chez les arbres forestiers car les obstacles sont nombreux (Muranty *et al.* 2014). Les QTL expliquent généralement une faible part de la variation phénotypique (<5 %) (Muranty *et al.* 2014). Par exemple, chez *Pinus radiata*, 8 QTL expliquent 14 % de la variation de la densité du bois juvénile, chaque QTL contribuant entre 0,8 et 3,6 % de la variation totale (Devey *et al.* 2004). Chez *Pinus pinaster*, 2 QTL ont été identifiés, expliquant 7 % et 12 % de la variation phénotypique de la hauteur de jeunes plants de 2 ans (Plomion *et al.* 1996a). De plus, l'association QTL-marqueurs dépend du fond génétique dans lequel elle a été mise en évidence. En effet, le déséquilibre de liaison est faible chez les arbres forestiers (Neale et Savolainen 2004). De plus, les arbres forestiers ont des génomes de très grande taille (par

exemple, le pin maritime a un génome estimé à 24,5 gigabases (Chagne *et al.* 2003)), ce qui complique encore la détection de QTL dans une population. Et même une fois mis en évidence, il est souvent compliqué de situer précisément un QTL sur un chromosome.

Par ailleurs, il a été montré que les QTL dépendent de l'âge et de l'environnement, ce qui constitue un autre obstacle pour la SAM.

➤ **Sélection génomique**

Contrairement à la SAM qui utilise un nombre restreint de marqueurs pour suivre un petit nombre de locus préalablement caractérisés (QTL ayant des effets importants sur le phénotype), la sélection génomique repose sur l'utilisation d'un grand nombre de marqueurs moléculaires. L'idée est de saturer le génome avec les marqueurs afin qu'ils soient en déséquilibre de liaison avec tous les QTL contrôlant le phénotype d'intérêt, même ceux à effets très faibles (Hayes et Goddard 2010). La connaissance *a priori* de la localisation et des effets des QTL ainsi que leur association avec les marqueurs n'est pas nécessaire. En effet, pour chaque marqueur, il s'agit d'estimer son effet individuel par modélisation statistique grâce à une population de calibration à la fois génotypée et phénotypée. La valeur génomique (GEBV) d'un individu est la somme des effets de tous ses marqueurs considérés (Meuwissen *et al.* 2001).

D'un point de vue statistique, l'évaluation génomique des individus repose sur l'analyse de l'information génomique d'un ensemble d'individus génotypés (variables explicatives) pour expliquer des performances observées (variables à expliquer) afin de construire le meilleur modèle prédictif. Une fois le meilleur modèle établi, on peut ensuite prédire les GEBV d'un ensemble d'individus candidats uniquement sur la base de leur génotype aux marqueurs. Dans une revue sur les bovins laitiers, Hayes *et al.* (2009) ont décrit les quatre facteurs principaux sur lesquels repose la précision des modèles de sélection génomique :

- i) l'héritabilité du caractère cible,
- ii) son architecture génétique (nombre et effets des QTL sous-jacents),
- iii) le niveau de déséquilibre de liaison entre les marqueurs et les QTL dans les populations de calibration et de validation,
- iv) la taille de la population de calibration, ainsi que le niveau d'apparentement entre les populations de calibration et de validation.

Les méthodes statistiques utilisées pour prédire les GEBV, qui diffèrent de par les hypothèses de distribution des effets des QTL, peuvent aussi influencer sur la précision de la GEBV (Heslot *et al.* 2015) mais dans une moindre mesure.

En sélection animale, la sélection génomique est déjà utilisée à grande échelle et s'est révélée très performante, notamment pour la sélection des taureaux (Boichard *et al.* 2012 ; Hayes *et al.* 2009 ; Schefers et Weigel 2012). Les jeunes taureaux sont génotypés avec une puce à ADN de référence de 54 000 SNPs répartis sur l'ensemble du génome bovin, et sont sélectionnés sur leur GEBV. La sélection génomique permet une très forte intensité de sélection : un très grand nombre de jeunes taureaux peut être génotypé, alors qu'un faible nombre peut être testé sur descendance à coût équivalent. De plus, elle diminue l'intervalle entre générations (gain de presque 3 ans (Schefers et Weigel 2012)) et offre un index de sélection très précis (VanRaden *et al.* 2009). Cela permet une meilleure réponse à la sélection et un gain génétique accru (il est doublé chez les bovins).



Chez les arbres forestiers, les programmes d'amélioration sont moins avancés qu'en sélection animale, mais l'utilisation de la sélection génomique en sélection forestière pourrait avoir de nombreux avantages (Grattapaglia et Resende 2011 ; Isik 2014). En effet, l'évaluation génétique sur descendance est long et très coûteux, l'intervalle de génération est long (plusieurs décennies) et les index de sélection peu précis.

Cependant, la sélection génomique n'est pas encore utilisée en sélection forestière (Isik 2014), même si de plus en plus d'études sont publiées à ce sujet (Beaulieu *et al.* 2014 ; Denis et Bouvet 2013 ; Grattapaglia et Resende 2011 ; Isik *et al.* 2016 ; Iwata *et al.* 2011 ; Resende *et al.* 2012a ; Resende *et al.* 2012b ; Zapata-Valenzuela *et al.* 2012). En effet, les obstacles sont encore nombreux. Les génomes des conifères sont de très grandes tailles (Mackay *et al.* 2012) (environ 24,5 Gb pour le pin maritime (Chagne *et al.* 2003)), un très grand nombre de marqueurs serait donc nécessaire pour saturer leurs génomes. De plus, ces espèces présentent un très faible déséquilibre de liaison (Neale et Savolainen 2004 ; Plomion *et al.* 2014), ce qui conduit à une faible transférabilité des modèles de prédiction génomique entre populations. Enfin, la taille efficace des populations des programmes d'amélioration forestiers est élevée (une grande diversité est conservée), ce qui diminue encore la probabilité d'association entre marqueurs et QTL.

Avant de pouvoir appliquer la sélection génomique dans leurs populations d'amélioration, les améliorateurs forestiers ont donc besoin de plus de marqueurs pour saturer les génomes des espèces de conifères sur lesquelles ils travaillent. Ils doivent aussi travailler sur des populations d'entraînement d'effectifs plus importants pour calibrer le modèle plus précisément et le rendre applicable à une population d'amélioration.

Même si les résultats obtenus en SAM et sélection génomique sur les arbres forestiers ne permettent pas aujourd'hui une sélection des individus à partir des données moléculaires, d'autres applications sont envisageables au sein des programmes d'amélioration forestiers, notamment grâce à la reconstitution de pedigree.

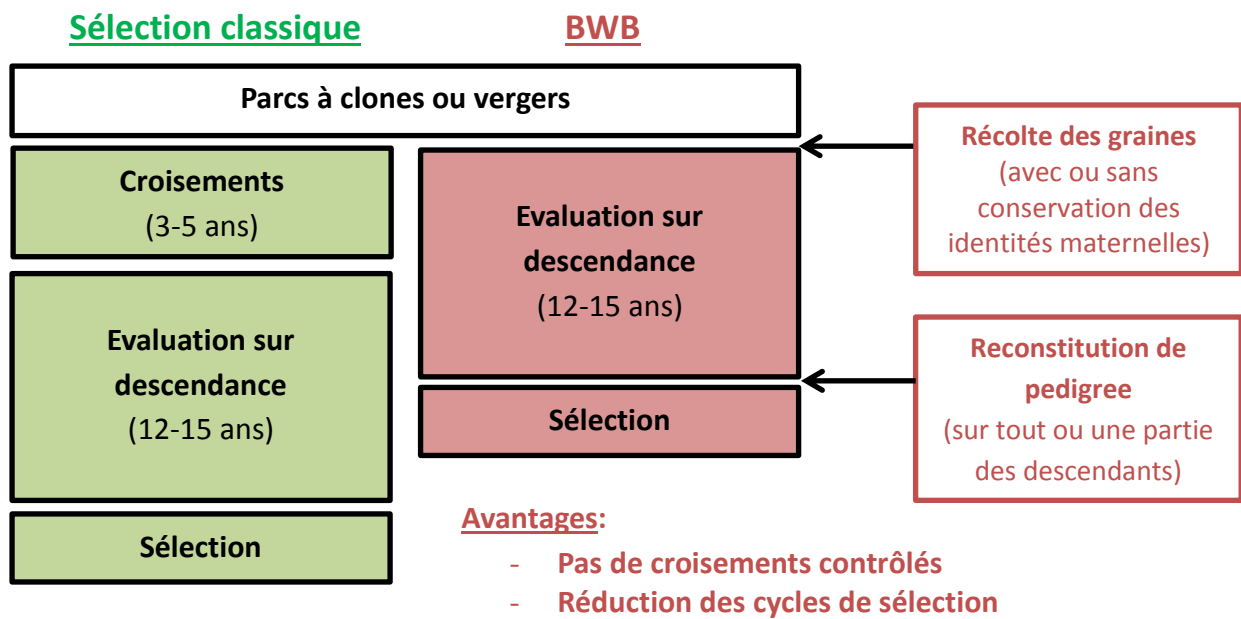
### **1.3.3 Des outils d'aide dans les programmes d'amélioration génétique dès aujourd'hui, grâce à la reconstitution de pedigree**

Le suivi du pedigree apparaît comme essentiel dans les programmes d'amélioration forestiers : il permet de contrôler l'apparement, l'évolution de la diversité génétique au cours des générations et une meilleure estimation des paramètres génétiques et des valeurs génétiques.

Cependant, le suivi du pedigree nécessite de réaliser des croisements contrôlés biparentaux, qui présentent de nombreux inconvénients :

- ils sont coûteux en temps et ressources,
- ils sont parfois difficiles à réaliser chez certaines espèces,
- ils limitent la taille efficace de la population d'amélioration,
- et ils ne sont pas exempts d'erreurs.

Figure 20 : Stratégie du « Breeding without breeding », proposée par El-Kassaby et Lstiburek (2009)



Les marqueurs moléculaires permettent de reconstituer le pedigree des individus (du moins lorsque l'on dispose des génotypes des parents potentiels). De nombreuses études récentes montrent son intérêt au sein de divers programmes d'amélioration forestiers (Bell *et al.* 2004 ; Bush *et al.* 2011 ; Doerksen et Herbinger 2008, 2010 ; El-Kassaby *et al.* 2011 ; El-Kassaby *et al.* 2010 ; El-Kassaby *et al.* 2015 ; Funda *et al.* 2014 ; Grattapaglia *et al.* 2014 ; Grattapaglia *et al.* 2004 ; Hansen et McKinney 2010 ; Hansen et Nielsen 2010 ; Kumar *et al.* 2007 ; Lai *et al.* 2010 ; Lambeth *et al.* 2001 ; Lstiburek *et al.* 2011 ; Moriguchi *et al.* 2011 ; Ng *et al.* 2012).

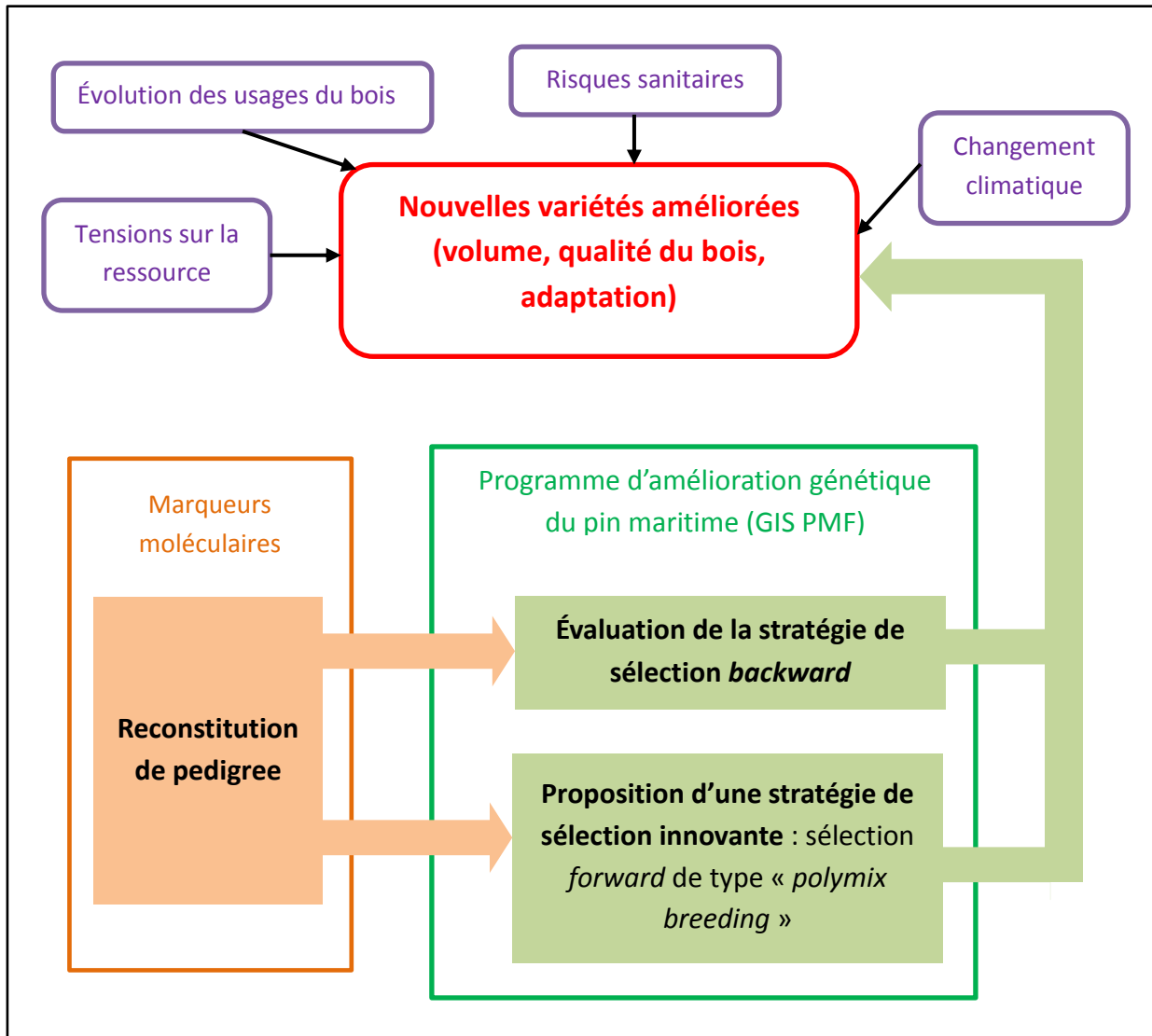
Une des applications principales de la reconstitution de pedigree à l'aide de marqueurs moléculaires est la simplification des croisements contrôlés biparentaux des programmes d'amélioration forestiers. Lambeth *et al.* (2001) ont proposé en premier le concept du « *polymix breeding with paternity analysis* », en remplaçant les croisements contrôlés biparentaux par un croisement polycross suivi d'une reconstitution de paternité. Cela présente un certain nombre d'avantages : simplicité logistique, moindre coût, maximisation du nombre de croisements, correction des erreurs d'identité des clones sélectionnés et de leur pedigree. El-Kassaby et Lstiburek (2009) ont élargi cette idée avec le concept « *Breeding without Breeding* » : ils proposent d'utiliser la pollinisation libre suivie d'une reconstitution de pedigree pour constituer la population d'amélioration. La durée du cycle d'amélioration du Douglas-Fir pourrait ainsi être réduite de 4 ans (figure 20), et on pourrait obtenir jusqu'à 85 % de la réponse génétique à la sélection qu'on aurait obtenue avec une stratégie d'amélioration classique (avec croisements contrôlés). Ce concept apparaît comme très intéressant pour débiter un programme d'amélioration, pour les espèces à faible importance économique, ou pour les programmes d'amélioration ayant peu de moyens.

La reconstitution de pedigree permet aussi de mieux estimer les paramètres génétiques et les valeurs génétiques lors de l'évaluation génétique sur descendance issue de croisements polycross ou pollinisation libre (Doerksen et Herbinger 2010 ; El-Kassaby *et al.* 2011 ; Gauzere *et al.* 2013 ; Hansen et McKinney 2010 ; Hansen et Nielsen 2010 ; Klapste *et al.* 2014 ; Korecky *et al.* 2013).

Enfin, la reconstitution de pedigree à l'aide de marqueurs moléculaires peut également être utilisée dans la gestion des vergers à graines, par exemple pour estimer les flux de gènes, la contamination du pollen (provenant de sources extérieures), pour étudier les modes de croisement, le taux d'autofécondation et la variation de la fertilité des individus mâles (El-Kassaby *et al.* 2010 ; Funda et El-Kassaby 2012 ; Funda *et al.* 2015 ; Hansen 2008 ; Lai *et al.* 2010 ; Machanska *et al.* 2013 ; Moriguchi *et al.* 2004 ; Moriguchi *et al.* 2011 ; Seido *et al.* 2000 ; Shimono *et al.* 2011 ; Torimaru *et al.* 2013) ou même pour vérifier les identités des clones.

Figure 21 : Problématique de la thèse

La création de variétés améliorées de pin maritime est essentielle pour relever les nouveaux défis et faire face aux contraintes de la filière forêt-bois (indiquées en violet). Les marqueurs moléculaires (en orange) constituent un outil à explorer pour optimiser le travail au sein du programme d'amélioration du pin maritime (en vert).



## 2. Problématique et objectifs de la thèse

Comme décrit dans la partie précédente, la filière forêt-bois en Aquitaine doit relever de nombreux défis tout en faisant face au changement climatique et aux risques sanitaires. En effet, il s'agit d'assurer la pérennité de la production de bois dans un contexte de tensions sur la ressource et d'évolution des usages du bois. Cette filière forêt-bois est très liée au pin maritime, qui est l'espèce majeure de la forêt des Landes de Gascogne (cf. [partie 1.1](#)). La production de nouvelles variétés génétiquement améliorées de pin maritime est l'un des leviers pour pouvoir faire face à ces défis. La création de variétés, alliant productivité et qualité tout en préservant l'adaptation environnementale, est l'objectif principal du programme d'amélioration génétique du pin maritime géré par le GIS « Pin Maritime du Futur » (cf. [partie 1.2](#)). Ces dernières années, grâce au développement des techniques de génotypage et de séquençage, les marqueurs moléculaires sont devenus plus facilement accessibles et sont de plus en plus utilisés en recherche forestière. Cependant, leur utilisation concrète au sein des programmes d'amélioration forestiers, y compris celui du pin maritime, reste encore très marginale (cf. [partie 1.3](#)).

Dans le but d'optimiser le travail au sein du programme d'amélioration du pin maritime, cette thèse s'intéresse à l'exploration de l'utilisation des marqueurs moléculaires afin de maximiser les gains génétiques et faire face aux défis de la filière forêt-bois en Aquitaine. Pour étudier cette problématique (figure 21), le génotypage est utilisé pour reconstituer *a posteriori* les pedigrees. D'une part, j'étudie si cela peut permettre d'optimiser la stratégie de sélection actuelle (stratégie de sélection *backward*), qui repose sur l'évaluation génétique des individus sur des tests de descendance polycross (dispositifs dans lesquels les arbres sont issus de croisements contrôlés entre une mère identifiée et un mélange de pollens appelé polymix). D'autre part, j'étudie la faisabilité de la mise en place d'une stratégie de sélection innovante (sélection *forward* basée sur la reconstitution de pedigree) permettant d'accélérer les cycles de sélection et de renouveler plus fréquemment les variétés.

Pour répondre à ces questions, trois objectifs ont été définis (figure 22) :

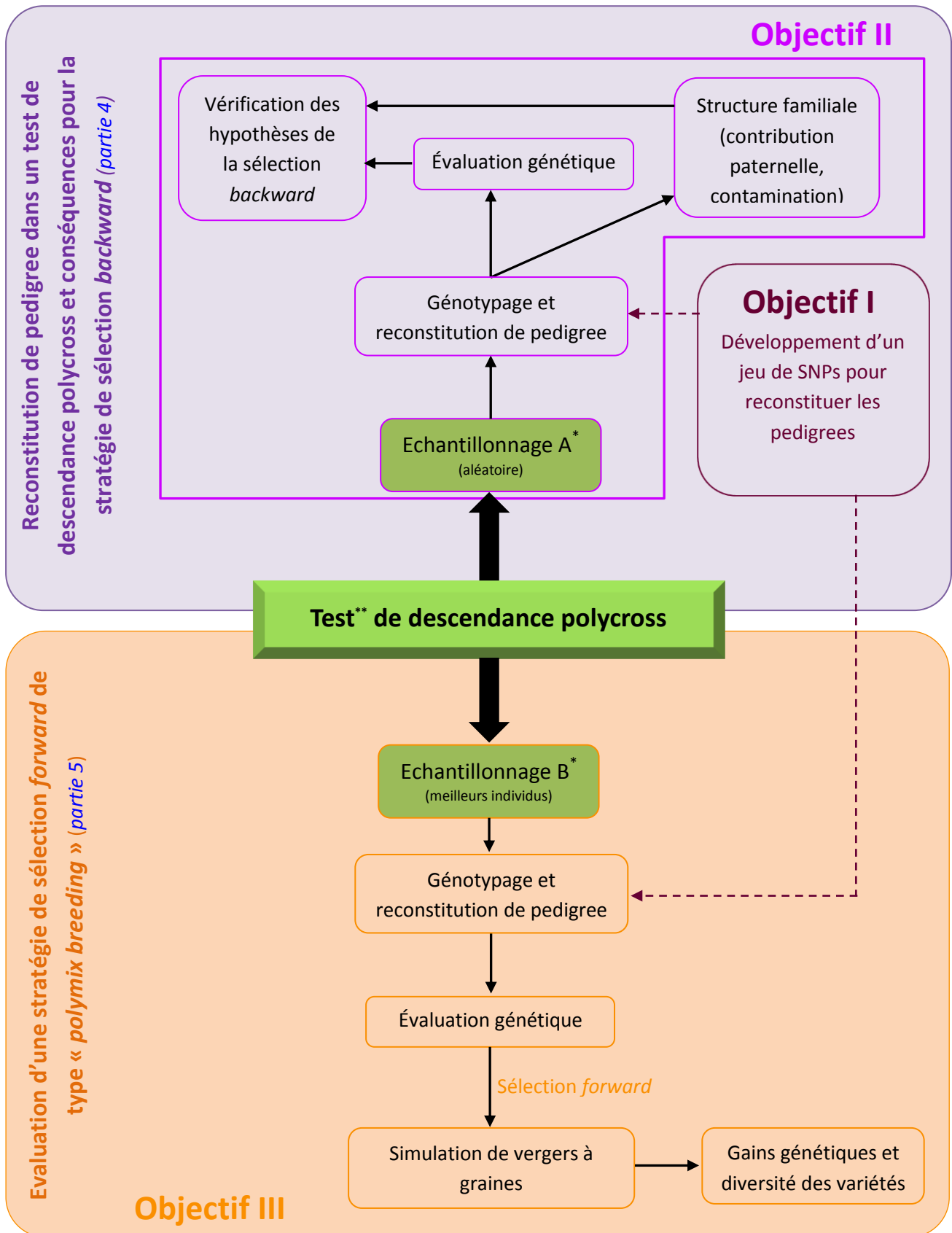
- Premièrement, j'ai développé un jeu de marqueurs moléculaires de type SNP pour permettre la reconstitution des pedigrees dans la population d'amélioration du pin maritime de provenance landaise. Ce travail constitue l'[objectif I](#) de la thèse.
- Deuxièmement, j'ai étudié si la connaissance d'un pedigree complet permettait d'optimiser la stratégie de sélection *backward*. J'ai donc reconstitué le pedigree de descendants échantillonnés aléatoirement dans un test de descendance polycross pour :
  - d'une part, tester les hypothèses (familles composées de demi-frères ; égale contribution des pères constituant le polymix) sur lesquelles repose la sélection *backward* classique (où seule l'identité de la mère est connue)
  - d'autre part, évaluer les conséquences sur l'estimation des paramètres génétiques et des valeurs génétiques des mères évaluées.

C'est l'[objectif II](#) de la thèse.



Figure 22 : Démarche méthodologique du projet de thèse

Les sections du manuscrit correspondantes sont indiquées en italique en **bleu foncé**.



\* L'échantillonnage est détaillé dans le matériel et méthode (cf. [partie 3.3](#))

\*\* Le dispositif étudié dans la thèse est le test de descendance polycross GIS3.3 (cf. [partie 3.1](#))

- Troisièmement, j'ai testé la faisabilité d'une stratégie de sélection innovante en réalisant une sélection *forward* de type « *polymix breeding* » au sein d'un test de descendance polycross (*i.e.* sélection *forward* parmi les descendants des tests polycross dont on aura au préalable reconstitué les pedigrees). Plusieurs vergers à graines sont constitués par simulation pour comparer les gains génétiques des variétés améliorées obtenues avec différentes stratégies de sélection. Ce travail constitue l'**objectif III** de la thèse.

Les résultats sont présentés et discutés dans les *parties 4* (qui traite des **objectifs I et II**) et *5* (qui traite de l'**objectif III**). Enfin, la *partie 6* décrit les principales applications et les perspectives qui découlent de cette thèse.

Une partie des travaux a été valorisée sous forme de deux publications :

- Article 1 (**objectifs I et II**), publié en 2015 dans *Tree Genetics and Genomes*.  
Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L (2015) **Paternity recovery in two maritime pine polycross mating designs and consequences for breeding**. *Tree Genet Genomes* 11:1-13. doi: 10.1007/s11295-015-0932-4
- Article 2 (**objectif III**), soumis en janvier 2016 à *Annals of Forest Science*.  
Vidal M, Plomion C, Raffin A, Harvengt L, Bouffier L (2016) **Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction**.

Figure 23 : Localisation du test de descendance polycross étudié, GIS3.3 à Cestas

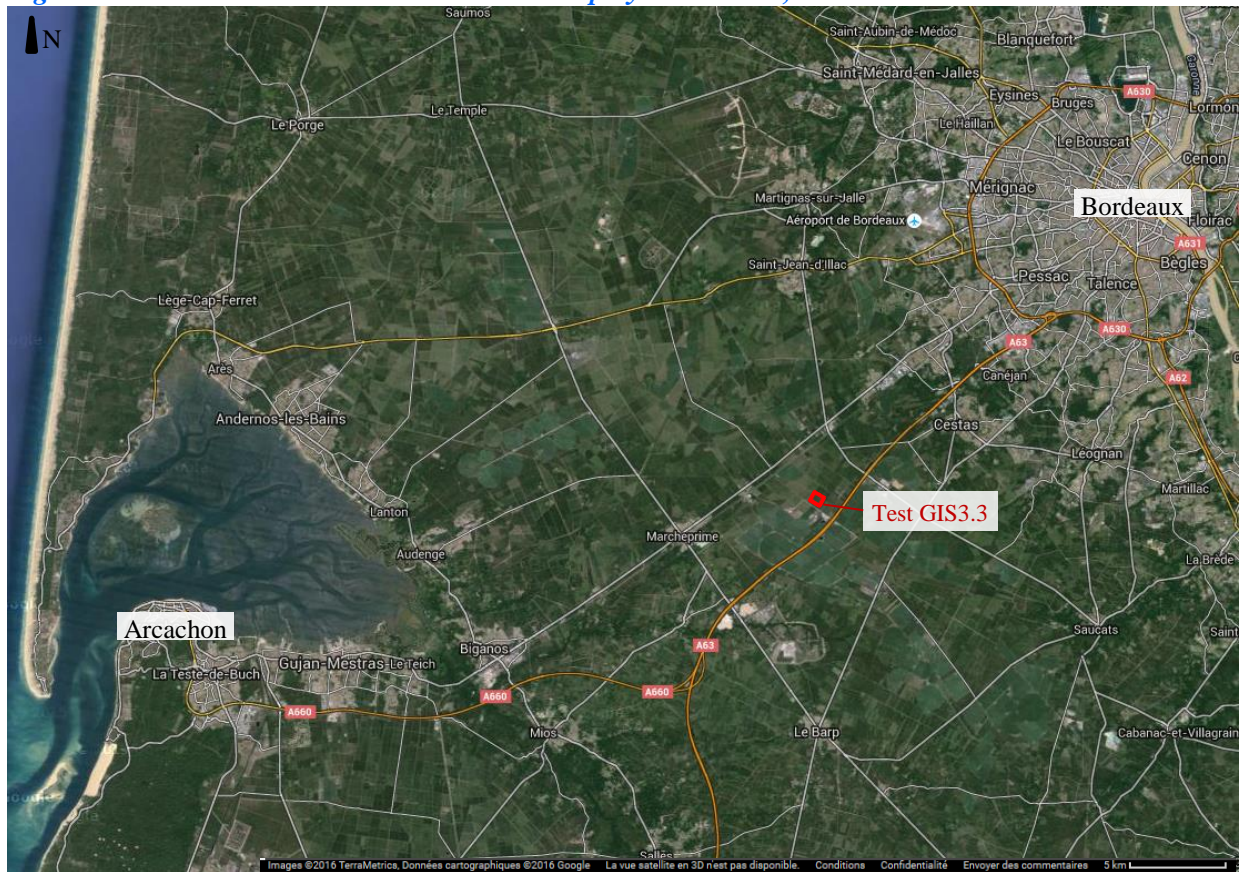


Figure 24 : Photographie du test de descendance GIS3.3, âgé de 17ans (prise de vue en 2013)



## 3. Matériel et méthodes

Cette partie décrit le matériel végétal et les principales méthodes utilisés dans ce travail de thèse pour répondre aux trois objectifs définis précédemment.

### 3.1 Test de descendance polycross étudié

L'étude a été réalisée sur l'un des tests de descendance polycross de pin maritime installés par le GIS PMF au sein du massif des Landes de Gascogne. Ce test, nommé GIS3.3, est constitué de 174 familles de demi-frères (6440 arbres) et a été installé sur 5 ha en 1996 en lande humide sur la commune de Cestas (33) (Figures 23 et 24). Un test de descendance polycross est un dispositif dans lequel chaque génotype à tester (ici de la génération G1) est pris comme mère et croisé avec un mélange de pollens (appelé polymix) d'environ 45 pères connus. Les performances des descendants installés sur ce site ont permis d'évaluer leurs mères G1 dans le cadre de la sélection *backward*.

Deux polymix (PMX) différents ont été utilisés pour la mise en place du test polycross GIS3.3: 98 et 76 mères ont été pollinisées respectivement avec le PMX\_A (mélange de pollens de 47 pères G1) et le PMX\_B (mélange de pollens de 43 pères G1). 8 mères ont été pollinisées avec les deux polymix. Aucun père n'est commun aux 2 polymix (Cf. [partie 4](#), « *material and methods* » de l'article 1 pour plus de détails).

### 3.2 Phénotypage des caractères d'intérêt

Pour rappel, le programme d'amélioration génétique a pour objectif principal l'augmentation de la productivité (*i.e.* croissance en volume) et l'amélioration de la rectitude basale du tronc.

Tous les individus du test de descendance GIS3.3 ont été mesurés à 8 ans pour l'écart à la verticalité (critère prédicteur de la rectitude basale du tronc) et à 12 ans pour la hauteur et la circonférence du tronc (critères prédicteurs de la croissance en volume). L'écart à la verticalité est mesuré avec une règle graduée munie d'un niveau à bulle et il représente la distance en centimètres entre la verticale et le tronc à 1,5 mètre du sol. La hauteur est mesurée en centimètres à l'aide d'un dendromètre à ultrasons (Vertex IV, Haglöf, Suède), constitué d'un transpondeur (cible placée sur le tronc) et d'un dendromètre calculant les distances et les angles. La circonférence du tronc est mesurée en centimètres à 1,30 mètre du sol à l'aide d'un mètre ruban.



### 3.3 Echantillonnage des descendants du test polycross

Pour des raisons de temps et de coût, tous les descendants G2 présents sur le test étudié GIS3.3 n'ont pas pu être génotypés. Deux échantillonnages distincts ont été réalisés :

- **Echantillonnage A**, pour répondre à l'objectif II (Etude de la stratégie de sélection *backward*) : **échantillonnage aléatoire** de 984 individus G2 (15 % du test de descendance), représentant 98 familles de demi-frères composées chacune de 9 à 12 descendants. 489 individus G2 (49 familles, avec 9 à 10 G2 par famille) sont issus de mères G1 pollinisées avec le PMX\_A ; 495 individus G2 (49 familles, avec 9 à 12 G2 par famille) sont issus de mères G1 pollinisées avec le PMX\_B (Cf. *partie 4*, « *material and methods* » de l'article 1).
- **Echantillonnage B**, pour répondre à l'objectif III (Evaluation d'une stratégie de sélection *forward*) : **échantillonnage des meilleurs individus** G2 du test (pré-sélection dans le cadre d'une stratégie *forward*). Les individus G2 du test sont classés selon un index génétique (BLUPs obtenus par une méta-analyse des dispositifs génétiques avec le logiciel Treeplan), prenant en compte le volume des arbres et la rectitude basale du tronc.

Deux types de présélections ont été étudiées (Cf. *partie 5*, « *material and methods* » de l'article 2) :

- Une sans contrainte en termes d'apparentements : les 153 meilleurs individus G2 du test polycross ont été échantillonnés,
- Une avec contrainte sur l'apparentement (2 individus par famille maternelle maximum) : les 2 meilleurs individus des 75 meilleures familles du test polycross ont été échantillonnés, soit 150 G2.

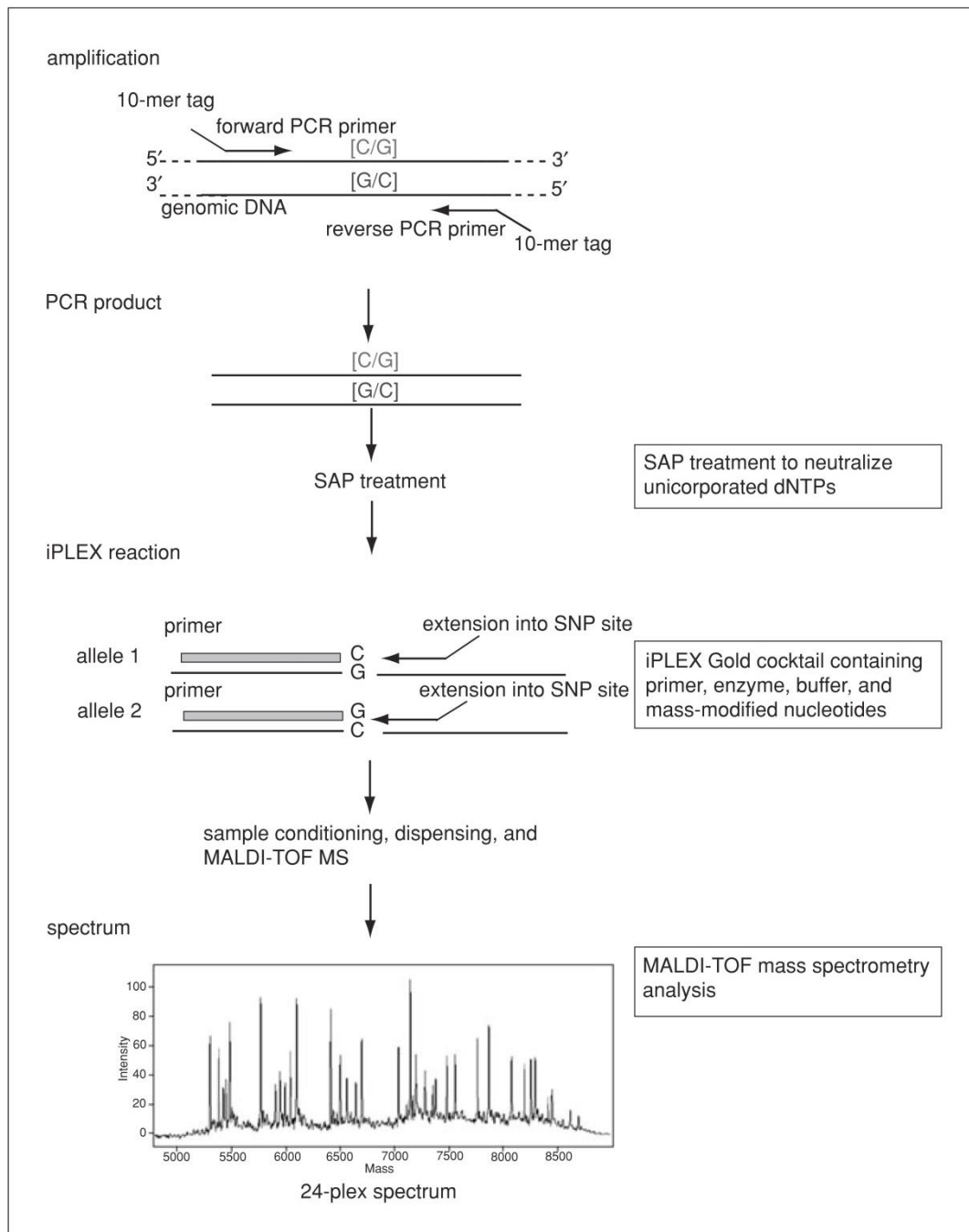
En pratique, l'échantillonnage sur le terrain consiste à récolter des aiguilles vertes sur les individus sélectionnés à l'aide d'un échenilloir (sécateur monté au bout d'une perche et actionné par l'intermédiaire d'une cordelette).

### 3.4 Génotypage

#### 3.4.1 Conception du jeu de SNPs

L'objectif I consiste à développer un jeu de marqueurs moléculaires SNPs permettant de reconstituer facilement les pedigrees dans la population d'amélioration du pin maritime. Pour cela, on s'est appuyé sur une base de données de SNPs polymorphes : préalablement à ce travail de thèse, 661 individus de génération G0 et G1 avaient été génotypés à l'aide d'une puce Infinium (Illumina, San Diego, USA) de 12 000 SNPs, développée par Chancerel *et al.* (2013). 2600 SNPs polymorphes ont ainsi été identifiés (Plomion *et al.* 2014). Afin de sélectionner les SNPs candidats pour la reconstitution de pedigree, nous avons d'abord retenus 169 marqueurs présentant une fréquence de l'allèle mineur (MAF : Minor Allele Frequency) supérieure à 0,45 afin de maximiser leur pouvoir de discrimination. Ensuite, le déséquilibre de liaison entre ces marqueurs a été calculé en prenant en compte l'apparentement entre les individus génotypés (package LDcorSV (Mangin *et al.* 2012) implémenté sous R). Seuls les SNPs non liés entre eux (*i.e.* avec  $r_v^2 < 0,3$ ) ont été retenus.

Figure 25 : Méthode de génotypage par réaction iplex (Sequenom), d'après Gabriel et al. (2009)



Parmi ces 121 SNPs d'intérêt non liés, nous avons créé deux multiplex de 40 SNPs chacun à l'aide du logiciel Assay Design Suite 1.0 Software (<https://www.mysequenom.com/Tools>). La liste des SNPs génotypés est disponible en ligne dans les données complémentaires<sup>8</sup> en annexe de l'article 1.

### 3.4.2 Génotypage iPLEX® Gold (Sequenom)

La technologie iPLEX® Gold proposée par Sequenom (San Diego, CA, USA) est fondée sur une PCR (amplification de la zone d'intérêt contenant le SNP) suivie d'une méthode d'extension d'une amorce par un nucléotide (méthode SBE, Single Base Extension) (Gabriel *et al.* 2009). Deux étapes constituent le typage des SNPs par spectrométrie de masse MALDI-TOF (Matrix-Assisted Laser Desorption/Ionisation-Time of Flight). D'abord, une étape de discrimination allélique correspond à la formation des produits réactionnels spécifiques à chaque allèle ou variant (à l'aide de ddNTPs, didésoxynucléotides). Ensuite, une étape de transfert des produits réactionnels sur une micropuce permet de détecter les produits issus de la discrimination allélique par spectrométrie de masse (Figure 25).

La conception des amorces PCR et SBE a été réalisée grâce au logiciel Assay Design Suite 1.0 Software (<https://www.mysequenom.com/Tools>). Le détail est disponible en ligne dans les données complémentaires<sup>8</sup> en annexe de l'article 1.

Le génotypage a été réalisé à la plateforme Génome Transcriptome (PGTB) à Cestas Pierroton (<http://www.pgtb.u-bordeaux2.fr/>).

### 3.5 Reconstitution de pedigree : méthode

Un grand nombre de méthodes statistiques permettent de réaliser les analyses de parenté (Jones *et al.* 2010 ; Meagher et Thompson 1986) , comme par exemple :

- La méthode d'exclusion : elle utilise les incompatibilités entre parents et descendants. Elle rejette l'hypothèse qu'un individu est issu d'un couple de parents s'il existe un seul mésappariement. Cette méthode n'est pas adaptée si le nombre de parents potentiels est très large et/ou si les marqueurs sont peu polymorphes (ce qui est le cas des SNPs, très souvent bialléliques). De plus, une simple erreur de génotypage suffit à rejeter un lien de parenté.
- La méthode d'allocation catégorique : elle repose sur une approche du maximum de vraisemblance afin de déterminer le parent le plus probable parmi une liste de parents potentiels. Un LOD score (*Likelihood Odds Ratio*) est calculé et indique la vraisemblance d'un couple d'individus (en analyse de parenté) à être les parents de l'individu que l'on cherche à assigner. Dans ce cas, un parent potentiel n'est pas exclu si le génotype attendu n'est pas correct sur quelques locus (mutation, erreur de génotypage, erreur de lecture).

C'est cette méthode d'allocation catégorique qui a été utilisée dans ce travail de thèse, à l'aide du logiciel Cervus 3.0 (Kalinowski *et al.* 2007 ; Marshall *et al.* 1998).

---

<sup>8</sup> [http://static-content.springer.com/esm/art%3A10.1007%2Fs11295-015-0932-4/MediaObjects/11295\\_2015\\_932\\_MOESM3\\_ESM.txt](http://static-content.springer.com/esm/art%3A10.1007%2Fs11295-015-0932-4/MediaObjects/11295_2015_932_MOESM3_ESM.txt)



Dans le test de descendance polycross étudié dans cette thèse, nous connaissons la mère théorique de chaque descendant. Les marqueurs ont été utilisés pour vérifier l'identité maternelle et retrouver le père parmi les pères potentiels constituant le polymix. Dans le cas d'une reconstitution de paternité, deux hypothèses sont testées : soit le père candidat est le vrai père de l'individu, soit il ne l'est pas. La vraisemblance de chacune de ces deux hypothèses est calculée selon la probabilité d'obtenir le génotype observé avec chaque hypothèse. Le rapport de vraisemblance (LOD score) est le logarithme décimal du ratio entre la probabilité qu'un individu (père potentiel) soit le vrai père et la probabilité que cet individu ne soit pas le vrai père. Plus ce rapport est élevé, plus le père candidat est probablement le vrai père.

Pour chaque marqueur SNP, le logiciel Cervus prend en compte les fréquences alléliques au sein des populations de descendants et des parents potentiels, ainsi que l'état homozygote ou hétérozygote du parent candidat. Ces informations permettent d'identifier le père candidat qui a le plus de probabilité d'être le vrai père. Des simulations permettent de générer 100 000 descendants aléatoirement (issus de la liste des parents potentiels) et de tenter de les réassigner à leurs parents. En fonction du nombre de descendants réassignés à leur « vrai » parent, un LOD score seuil est fixé selon l'intervalle de confiance demandé. Ainsi, on cherche à minimiser les deux erreurs suivantes : considérer qu'un descendant n'a pas de père identifié parmi les pères candidats alors qu'il en a un, et considérer qu'un père candidat est le père d'un descendant alors qu'il ne l'est pas.

L'identité de la mère théorique est confirmée lorsque le LOD score est positif (et avec au maximum un seul mésappariement entre le descendant et sa mère supposée). Pour l'analyse de paternité, les pères potentiels sont les individus constituant les polymix. On considère un taux de contamination de 10 %, c'est-à-dire que 10 % des descendants ont un père qui n'appartient pas au polymix. Ce paramètre réduit la puissance d'assignation, et donc l'analyse est encore plus conservatrice. Au vu du pouvoir de discrimination de notre jeu de marqueurs SNPs utilisés dans cette étude, l'intervalle de confiance est fixé à 99 %.

### 3.6 Evaluation génétique

Dans la *partie 4*, pour répondre à l'**objectif II**, on souhaite observer l'impact de la reconstitution de pedigree sur l'estimation des valeurs génétiques maternelles (dans le cas d'une évaluation sur descendance). La résolution des équations du modèle mixte, permettant d'estimer les paramètres génétiques et les valeurs génétiques pour la circonférence à 12 ans et l'écart à la verticalité à 8 ans, a été réalisée avec ASReml version 3.0 (Gilmour *et al.* 2009). Seules les données phénotypiques des individus G2 du test étudié (test de descendance GIS3.3 à Cestas) ont été prises en compte dans la résolution des équations. Deux modèles ont été appliqués : le modèle « pedigree partiel », en utilisant la matrice de parenté **A** réalisée avec information de la mère, et le modèle « pedigree complet », en utilisant la matrice de parenté **A** réalisée avec information des deux parents (obtenue après reconstitution de pedigree). Ces analyses sont détaillées dans le « *Material and methods* » de l'article 1 (*partie 4*).

Dans la *partie 5*, pour répondre à l'**objectif III**, on se place dans une optique d'application de sélection *forward* au sein d'un test de descendance. L'évaluation génétique a été réalisée avec le logiciel Treeplan pour valoriser l'ensemble des données du programme d'amélioration. Il s'agit donc d'une analyse multi-sites et multi-caractères.

## 4. Reconstitution de pedigree dans un test de descendance polycross et conséquences pour la stratégie de sélection *backward*

Cette *partie 4* traite des *objectifs I et II* de la thèse.

- L'*objectif I* consistait à développer un jeu de marqueurs moléculaires SNP permettant de reconstituer facilement et à un coût accessible les pedigrees dans la population d'amélioration du pin maritime de provenance landaise.
- L'*objectif II* consistait à utiliser la reconstitution de pedigree pour tester les hypothèses sur lesquelles reposent la sélection *backward*, et d'en évaluer les conséquences sur l'estimation des paramètres génétiques et des valeurs génétiques des mères évaluées.

Cette partie a fait l'objet d'un article intitulé « *Paternity recovery in two maritime pine polycross mating designs and consequences for breeding* », publié dans la revue *Tree Genetics and Genomes* en septembre 2015.

D'autre part, elle a fait l'objet d'une communication orale au congrès international IUFRO « *Tree Biotechnology conference* » en juin 2015 (Florence, Italy) et d'un poster au congrès international IUFRO « *Forest Genetics* » en juillet 2013 (Whistler, BC, Canada).

### 4.1 Résumé de l'article

L'amélioration génétique du pin maritime repose sur une stratégie de sélection récurrente. Cette stratégie consiste en une succession de cycles associant une phase d'évaluation génétique, une phase de sélection et une phase de croisement. La phase d'évaluation génétique consiste à évaluer les individus sur les performances de leurs descendants. Ainsi, de nombreux tests de descendance polycross ont été installés pour évaluer et classer les individus utilisés en tant que mères (dans le cadre de la stratégie de sélection *backward*).

Comme l'identité des pères des descendants constituant le test polycross est inconnue, l'estimation des paramètres génétiques et des valeurs génétiques repose sur l'hypothèse que les descendants d'une même mère constituent une famille de demi-frères et qu'il y a égale contribution des pères du polymix au sein de la descendance. L'objectif de cette étude était de tester ces hypothèses et d'en évaluer les conséquences sur l'estimation des paramètres génétiques et des valeurs génétiques dans le cadre d'une sélection *backward*.

Pour cela, un échantillonnage aléatoire (*échantillonnage A*) a été réalisé dans l'un des tests de descendance polycross installés par le GIS PMF : 984 individus G2 (soit 15 % du test

#### 4. Reconstitution de pedigree dans un test de descendance polycross et conséquences pour la stratégie de sélection backward

de descendance étudié), représentant 98 familles de demi-frères composées chacune de 9 à 12 descendants, ont été génotypés à l'aide de 80 marqueurs moléculaires de type SNP. Les parents potentiels (mères évaluées G1 et pères constituant les polymix) ont également été génotypés.

Après analyse de parenté avec un niveau de confiance de 99 %, le père (appartenant au polymix) est identifié pour 89 % des descendants G2 génotypés. On a montré que certains pères du polymix contribuent plus que d'autres (entre 1 et 26 descendants par père, avec une contribution moyenne de 9,3 descendants) mais l'écart à l'hypothèse d'égale contribution des pères est dû à seulement quelques pères (contribuant plus ou moins qu'attendu). Au sein des 98 familles étudiées, on compte 32 familles constituées uniquement de plein-frères et 8 familles avec plus de 3 plein-frères, parmi les 10 descendants génotypés dans chaque famille. Le coefficient d'apparentement moyen intra-famille est de 0,26 (contre 0,25 dans le cas de famille de demi-frères), donc l'hypothèse des familles de demi-frères est très majoritairement respectée.

Les valeurs génétiques du diamètre (à 12 ans) et de l'écart à la verticalité (à 8 ans) des mères G1 (évaluées dans le test polycross) estimées avec l'information du pedigree partiel (mère théorique) ou du pedigree complet (après reconstitution de pedigree) sont fortement corrélées (coefficient de corrélation de Pearson de 0,96 et 0,97 pour l'écart à la verticalité et le diamètre, respectivement). On en conclut que l'évaluation génétique des mères G1 et leur classement dans le cadre de la sélection *backward* est fiable même si le pedigree complet n'est pas connu.

Enfin, la reconstitution de pedigree dans un test de descendance polycross offre la possibilité de réaliser une sélection *forward*, c'est-à-dire que les descendants G2 génotypés avec un pedigree complet peuvent alors être sélectionnés pour constituer la population d'amélioration de la génération suivante et/ou pour établir une population de production.

## 4.2 Article 1



## Paternity recovery in two maritime pine polycross mating designs and consequences for breeding

Marjorie Vidal<sup>1,2,3</sup> · Christophe Plomion<sup>1,2</sup> · Luc Harvengt<sup>3</sup> · Annie Raffin<sup>1,2</sup> ·  
Christophe Boury<sup>1,2</sup> · Laurent Bouffier<sup>1,2</sup>

Received: 26 June 2015 / Revised: 4 September 2015 / Accepted: 9 September 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** Polycross mating systems are widely used in forest tree breeding for genetic testing. Backward selection based on polycross testing assumes equal male reproductive success and true half-sib progeny. The main objectives of this study were, firstly, to investigate the departure from these assumptions in a maritime pine polycross trial and, secondly, to evaluate the consequences for heritability and breeding values estimations. A total of 984 offspring from 98 half-sib families was genotyped with single nucleotide polymorphism markers to recover the full pedigree. Paternity was assigned successfully for 89 % of the offspring at a 99 % confidence level. We thus concluded there was an 11 % pollen contamination rate, assuming contamination when no genotype from the polymix composition could be identified as a father. The paternal contribution to the offspring varied among the males, but the departure from half-sib assumption was moderate since the average genetic correlation within the family was 0.26. Heritability and breeding values for girth at breast height and stem sweep were estimated using individual-tree mixed

models with either partial or full pedigree information. The results highlighted a minor bias in heritability estimation due to unknown paternity, as well as a high correlation for estimated breeding values between the partial and full pedigree models, suggesting that the genetic merit of the parental generation for backward selection was adequately predicted using the partial pedigree model. Finally, pedigree recovery was also discussed in a perspective of forward selection.

**Keywords** Pedigree reconstruction · SNP markers · Paternal reproductive success · Pollen contamination · Breeding strategies · *Pinus pinaster* Ait.

### Introduction

Tree breeding generally follows a classical recurrent selection scheme, which is characterized by repetitive cycles of breeding, testing, and selection (White 1987; Zobel and Talbert 1984). In conifers, selection of best individuals for either further breeding or inclusion in a production population (seed orchard) is commonly performed based on additive genetic values since open-pollinated seed orchards are generally used for deployment. The breeding value of a genotype is defined as the sum of the average effects of its genes and is estimated here with the performances of its progeny. Thus, large numbers of polycross trials were established to assess breeding values and to rank female parents for backward selection (Burdon and Shelbourne 1971). Polycross mating is done by mixing pollen from several males and applying the resulting polymix to isolated female flowers. As the male parent of each individual offspring is unknown, estimates of genetic parameters and breeding values are based on the assumption that all progeny of a female are true half-sibs (Squillace 1974) and that an equal contribution of male parents takes place

Communicated by: J. Beaulieu

This article is part of the Topical Collection on *Breeding*

**Electronic supplementary material** The online version of this article (doi:10.1007/s11295-015-0932-4) contains supplementary material, which is available to authorized users.

✉ Laurent Bouffier  
laurent.bouffier@pierroton.inra.fr

<sup>1</sup> INRA, BIOGECO, UMR 1202, 33610 Cestas, France

<sup>2</sup> University Bordeaux, BIOGECO, UMR 1202, 33400 Talence, France

<sup>3</sup> FCBA, Biotechnology and Advanced Silviculture Department, Genetics & Biotechnology team, 33610 Cestas, France

(Moriguchi et al. 2009). Departure from these assumptions could result in overestimation of both additive genetic variance and prediction of genetic gain, as relatedness would be underestimated.

Thanks to the availability of affordable and reliable DNA markers, coupled with powerful pedigree reconstruction methods, incomplete pedigree trials can be converted into complete pedigree trials. Most pedigree recovery studies have shown that paternal contributions in controlled polycross trials were unequal (Moriguchi et al. 2009); this has been confirmed for *Pseudotsuga menziesii* (Nakamura and Wheeler 1992), *Betula pendula* (Pasonen et al. 1999), *Picea abies* (Aronen et al. 2002), *Populus* spp. (Wheeler et al. 2006), *Pinus radiata* (Kumar et al. 2007), and *Picea rubens* Sarg (Doerksen and Herbinger 2008). Equal paternal contribution was found in the progeny of a polycross (with nine pollen donors) in *Pinus taeda* (Wiselogle and Vanbuijtenen 1988). A differential reproductive success between males does not necessarily lead to a substantial increase of the genetic correlation among offspring within the same family (Doerksen and Herbinger 2010; Hansen and Nielsen 2010) or to an overestimation of additive genetic variance (Doerksen and Herbinger 2010; Hallingback and Jansson 2013). Recent studies have investigated the impact of unequal paternal contributions and/or selfing on genetic parameters and estimated breeding values (EBVs) by genotyping offspring originating from open-pollinated seed orchards (El-Kassaby et al. 2011; Gaspar et al. 2009; Hallingback and Jansson 2013; Hansen and McKinney 2010; Hansen and Nielsen 2010; Klapste et al. 2014; Korecky et al. 2013) or from polycross mating designs (Doerksen and Herbinger 2010; Kumar et al. 2007). These authors observed that individual heritability estimated without paternity information could be larger or smaller, in comparison with their respective estimates including paternity information.

In the French maritime pine (*Pinus pinaster* Ait.) breeding program, selection of the production population is based on polycross testing (backward selection) on the one hand, and on the other, selection of the following breeding generation is based on full-sib family testing (forward selection).

For this species, representing the most commonly planted forest tree in France, large genomic resources have been recently published including a unigene set (Canales et al. 2014) and a catalog of single nucleotide polymorphism (SNP) markers (Chancerel et al. 2013). These molecular markers constitute reliable tools to recover paternal identity in polycross trials. In this context, the main objective of this study was to investigate, for the first time in maritime pine, departure from assumptions of backward selection based on polycross trials. To this end, we firstly recovered paternity identities of 970 successfully genotyped offspring (out of 984 offspring sampled) obtained from two different polymixes (PMX) using SNP markers, we then assessed the

pollen contamination rate (*i.e.* proportion of pollen donors not belonging to the PMX composition), studied the deviation from equal paternal contribution, and tested whether the offspring of female parents were true half-sibs; finally, we studied the bias in the estimation of heritability and breeding values due to unknown paternal identity.

## Materials and methods

### Plant material and mating design

Six polycross trial series were established from 1994 to 2002 in southwestern France to predict first-generation (G1) parental breeding values of the maritime pine breeding program. G1 trees are selected progeny from G0 trees, whereas G0 trees were mass-selected from the Landes provenance and constitute the base population of the breeding program (Illy 1966). Each series was composed of three polycross trials established on contrasted sites for water availability (humid, semi-humid, and dry) with the same half-sib families. One trial of the third series was considered in this study (a humid site located 44° 42' 32" N / 0° 46' 8" W). It consisted of a 5-ha polycross trial established in 1996 with 174 half-sib (HS) families plus five checklots, planted in a randomized block design, totaling 6440 trees (35 complete blocks with one-tree plot per family and two trees per checklot in each block). Two different pollen mixes were used: 98 seed parents pollinated with polymix A (PMX\_A, 47 G1 pollen donors) and 76 seed parents pollinated with polymix B (PMX\_B, 43 G1 pollen donors). Among those G1 seed parents, 8 were pollinated with both PMX\_A and PMX\_B (families from identical seed parents but distinct PMX were considered as being different). There were no pollen donors in common between PMX\_A and PMX\_B.

Control-pollination was realized in April 1994 as described below. Isolated pollen cones were collected from the 47 (PMX\_A) and 43 (PMX\_B) pollen parents and processed following standard pollen extraction methods (pollen dried at 30 °C with low relative humidity during 48–72 h). Pollens were kept at 5 °C until they were used. Both polymixes (PMX\_A and PMX\_B) consisted of equal volumes of their components. On each seed parent tree, seed cones were isolated with pollination bags approximately 10–15 days before receptivity (scales still closed). The isolated cones were pollinated with one polymix three times during their maximum receptivity stages using a squeeze bulb syringe system to spray the pollen (holes in bags resulting from the pollen injection were plugged to prevent pollen contamination). Pollination bags were removed at the post-receptivity stage, generally 10–15 days after the last pollination and a durable identification tag was fixed. The cones resulting from the controlled pollinations were harvested 19 months later and then dried at 30 °C during 3 months before extraction.

The offspring of both polycrosses (G2 trees) were phenotyped: tree girth at breast height was measured at the age of 12 years, and sweep (stem deviation from verticality at 1.5 m expressed in cm) was measured at the age of 8 years.

### Sampling

Two successive samplings were carried out:

- i. In spring 2012, young green needles of 489 G2 individuals (49 families × 9–10 individuals/family) from HS families obtained with PMX\_A were collected (Subset\_A).
- ii. In spring 2013, young green needles of 495 G2 individuals (49 families × 9–12 individuals/family) from HS families obtained with PMX\_B were collected (Subset\_B).

A total of 984 G2 genotypes (15 % of the progeny trial) was thus sampled, representing 98 HS families with 9–12 randomly sampled trees per family. Moreover, needles of the putative parental genotypes (seed and pollen donors) were collected (considering two ramets from two different clonal archives when available). Subset\_A came from 49 theoretical females and 47 males (PMX\_A). However, one pollen donor belonging to PMX\_A could not be genotyped (since it was missing in the clonal archives). Four genotypes were used both as males and females, so 91 parental genotypes were collected to analyze Subset\_A. Subset\_B came from 49 theoretical females and 43 males (PMX\_B). As 10 genotypes were used as males and females, 82 parental genotypes were collected to analyze Subset\_B.

### DNA extraction and fingerprinting

Needles were kept in a cool box during sampling and then stored at –80 °C. DNA extractions were carried out on fine powder of needle tissues using an Invisorb® DNA Plant HTS 96 Kit (Stratec Molecular, Berlin, Germany), in compliance with the manufacturer's instructions and quantified with a NanoDrop microvolume spectrophotometer (Thermo Fisher Scientific Inc., Waltham, CA, USA).

Genotyping A: From the 12 k Infinium SNP-array (Illumina, San Diego, USA) developed by Chancerel et al. (2013), 2600 polymorphic SNPs were available from a previous screening of 661 G0 and G1 trees (Plomion et al. 2014). A first filter was used to retain 150 SNPs with minor allele frequency (MAF) greater than 0.45 to maximize the discrimination power of the marker. Assay Design Suite 1.0 software (<https://www.mysequenom.com/Tools>) from Sequenom was finally used to automatically design two sets of iPLEX single base extension primers comprising 32 and 33 SNPs. Subset\_A was genotyped with these 65 SNPs using the Sequenom MassARRAY iPLEX Gold assays (Sequenom,

San Diego, CA, USA) in the genotyping and sequencing facility of Bordeaux, France (<http://www.pgtb.u-bordeaux2.fr/>).

Genotyping B: Subset\_B was genotyped with an optimized set of SNP markers. The most informative and easily read SNPs (39 out of 65) from the above-mentioned genotyping assay were retained and completed with an additional set of 41 SNPs, considering not only the MAF (>0.45) but also the extent linkage disequilibrium, a criteria that was not taken into account in Subset\_A. Pairwise linkage disequilibrium among candidate SNPs was calculated with LDcorSV package (Mangin et al. 2012) implemented in R (R Core Team (2013)) that corrects for relatedness, and only unlinked SNPs ( $r_v^2 < 0.3$ ) were kept. Thus, two new multiplexes of 40 SNPs each were designed to genotype Subset\_B (cf. supplemental information for SNPs details).

### Parentage assignment

Likelihood inference methodology was carried out with Cervus 3.0 (Kalinowski et al. 2007; Marshall et al. 1998), both to check the female (seed parent) identity and to recover the paternal (pollen donor within each PMX) identity of G2 trees. Cervus was run assuming a 0.1 % genotyping error rate (estimation based on repeated genotyping of 50 G0 and 136 G1 individuals). G2 trees with less than 42 loci successfully genotyped in both subsets were excluded from the parentage analysis.

The female parent was confirmed when the LOD score was positive, and only one mismatched allele was allowed for each offspring and its supposed female parent.

For paternity recovery, we considered that 90 % of the pollen donors had been sampled (we assumed a 10 % contamination rate, namely the presence of pollen donors exterior to PMXs). The delta score (*i.e.* the difference in LOD scores of the two most likely candidate parents) was used as a criterion for assignment of paternity at 99 % level of confidence. Critical values of delta scores were based on simulations of 100,000 offspring. One mismatched allele was allowed between a given offspring and its parent. The Cervus selfing option was used, as some potential pollen donors were also seed parents.

The effective number of males  $N_{em}$  (expected and observed values) was calculated for each subset using the following formula (Kimura and Crow 1963):

$$N_{em} = \frac{N_m k_m - 1}{k_m - 1 + \frac{V_m}{k_m}} \quad (1)$$

where  $N_m$  is the number of males,  $k_m$  the mean number of progeny produced by males, and  $V_m$  the variance of the number of progeny per male.

### Relatedness within and between families

Relatedness within and between families was analyzed using the additive genetic relationship coefficients from the additive genetic relationship matrix (A). The A matrix (or numerator relationship matrix) stems from the pedigree and contains diagonal elements equal to  $1+F_i$ , where  $F_i$  is the inbreeding coefficient of individual  $i$ . The off-diagonal elements ( $A_{ij}$ ) specify the relationship coefficient between individuals  $i$  and  $j$  ( $A_{ij}=2\theta_{ij}$ , where  $\theta_{ij}$  is the coefficient of coancestry). The coefficient of coancestry  $\theta_{ij}$  describes the relatedness between pairs of individuals that is equivalent to the probability that two gametes taken at random (one from each) carry alleles that are identical by descent (e.g., for half-sibs  $2\theta_{ij}=0.25$ , for full-sibs  $2\theta_{ij}=0.5$ , when parents are non-inbred and non-related) (Lynch and Walsh 1998).

The A matrix was calculated with the synbreed package in R (R Development Core Team 2013), taking into account full pedigree of G1 parents (*i.e.* considering their G0 parents) and partial or full pedigrees (depending on the model considered) of G2 progeny. The average relatedness of a family was estimated with the additive relationship coefficient, which is the mean of the off diagonal elements of the genotyped progeny of this family.

### Variance–component estimation and BLUP breeding value prediction

Mixed models were solved by restricted maximum likelihood implemented in ASReml version 3.0 (Gilmour et al. 2009) to estimate variance components and best linear unbiased predictors (BLUP) for girth at 12 years of age and sweep at 8 years of age.

The analyses were based on individual-tree linear mixed models of the form given in Eq. 2:

$$y = Xb + Za + e \quad (2)$$

where  $y$  is the vector of individual tree observations on each trait,  $b$  is a vector of fixed-effect estimates (block effect),  $a$  is a vector of random additive genetic effects, and  $e$  is a vector of random residual effects.  $X$  and  $Z$  are incidence matrices for fixed and random model terms.

The additive and error effects were assumed uncorrelated with zero means and normally distributed as  $a \sim N(0, A\sigma_a)$  and  $e \sim N(0, I\sigma_e)$ , where  $I$  is an identity matrix and  $A$  was the additive relationship matrix.

Narrow-sense heritability ( $h^2$ ) was calculated for each trait as:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (3)$$

where  $\sigma_a^2$  is the additive genetic variance and  $\sigma_e^2$  is the residual variance.

The general mixed model was fitted to the quantitative data of the successfully genotyped G2 trees (Subset\_A and Subset\_B together), including different pedigree files according to the model considered.

Here, the term “model” was used to differentiate the varying degrees of pedigree information (with and without paternity identity), which were fit to the same individual-tree model described in Eq. 2. In the partial pedigree (PP) model, only the theoretical seed donors of the families were known. In the full pedigree (FP) model, the pedigree file contained both parents’ identities (from paternity recovery).

In addition, for both models, the pedigree files in ASReml always contained the identity of the grandparents (G0), when available.

Best linear unbiased predictions, named estimated breeding values (EBVs), were estimated for the parental clones as well as for their progeny. Female breeding values from both models were compared by Pearson product–moment correlation and Spearman rank-order correlation to examine whether the complete pedigree information caused significant changes in EBVs as well as ranking that would affect selection and predicted gains.

The accuracy of predicted breeding values was calculated as:

$$r = \sqrt{1 - \frac{PEV}{(1 + F)\sigma_a^2}} \quad (4)$$

where PEV is the prediction error variance,  $F$  is the inbreeding coefficient, and  $\sigma_a^2$  is the additive genetic variance (El-Kassaby et al. 2011).

## Results

### SNP genotyping

Four hundred eighty nine offspring (Subset\_A) and their respective 91 G1 parents were genotyped with 65 SNP markers, and 495 offspring (Subset\_B) and their respective 82 G1 parents were genotyped with 80 SNP markers.

A SNP marker was considered unreliable (and therefore excluded from further analyses) when the proportion of individuals displaying assay failure was greater than 30 % in the Sequenom output data. Overall, 56 and 63 SNPs were finally retained for Subset\_A and Subset\_B, respectively. Their positions on the maritime pine linkage map from Chancerel et al. (2013) are provided as supplemental information.

When all loci were considered, the combined power of exclusion was higher than 0.99 for both sets of markers.

Moreover, the LD between the 63 SNPs of the optimized markers set used to genotype the Subset\_B was very low ( $r^2$  mean= $2.5 \cdot 10^{-3}$ , min= $1.9 \cdot 10^{-9}$ , max=0.24).

It should be noted that most (970 out of 984) offspring were genotyped with at least 42 SNPs. The 14 G2 trees below this threshold were simply removed from further analysis (Table 1). Similarly, all G1 parents (except one pollen donor from PMX\_A which was not available in clonal archives) were successfully genotyped with at least 42 SNPs. A fraction of males (62 out of the 90 pollen donors) were genotyped twice (by collecting needles on two different ramets from two clonal archives) and compared: all the identities were confirmed.

### Paternity recovery

The maternal identities were the first ones we checked. A total of 41 offspring (out of the 970 analyzed) did not conform to their nominal seed donor genotypes. Two types of pedigree errors were highlighted: (i) when the error was related to only one G2 offspring in a given family, it was likely that needles of that offspring were collected on an erroneous genotype (this was the case for 11 offspring); (ii) when the error was related to all offspring of a given family, mishandled maternal identities were highly probable (this was the case for 3 seed parents and their 10 respective offspring).

All these 41 offspring, identified as maternal mismatches, were excluded from paternity analyses, and thus, the paternity recovery was performed on the 929 remaining progeny (460 in Subset\_A and 469 in Subset\_B, cf. Table 1). Paternity was successfully recovered for 828 offspring: 87 and 91.3 % were assigned to a pollen donor member of the PMX (at a confidence of 0.99) in Subset\_A and Subset\_B, respectively.

**Table 1** Descriptive statistics about paternity recovery (with CERVUS) in two maritime pine polycross breeding populations

Parameter	Subset_A	Subset_B
Total number of sampled HS families	49	49
HS family size	9–10	9–12
Number of offspring sampled	489	495
Number of reliable SNPs genotyped	56	63
Number of offspring successfully genotyped	478	492
Number of offspring with seed donor confirmed	460	469
Number of offspring sired by a PMX pollen donor	400	428
Number of offspring sired by foreign pollen	60	41
Pollen contamination rate	0.13	0.087
Number of contributing males	46	43
Mean male contribution (nb of offspring)	8.70	9.95
Variance in male contribution	26.53	28.76

Subset\_A and Subset\_B correspond to two different PMXs

### Pollen contamination rate in the polymix mating

In this study, we considered the pollen contamination rate as the proportion of offspring which could not be assigned to a known pollen parent of either PMX with a statistical support >99 %. Overall, 60 (13 %) and 41 (8.7 %) offspring in Subset\_A and Subset\_B, respectively, were clearly fathered by outside pollen (Table 1). The mean pollen contamination rate over the whole sampling was therefore 10.9 %.

The contamination rate varied from 0 to 70 %, depending on the family (with 9 to 12 offspring genotyped per family). In Subset\_A, 17 families were not contaminated by foreign pollen, 25 families had 1 or 2 offspring sired by foreign pollen parents, and 6 families presented 3 or more offspring sired by foreign pollen parents. In Subset\_B, 18 families were not contaminated by foreign pollen, 28 families exhibited 1 or 2 offspring sired by foreign pollen parents and only 1 family had 3 offspring sired by foreign pollen parents.

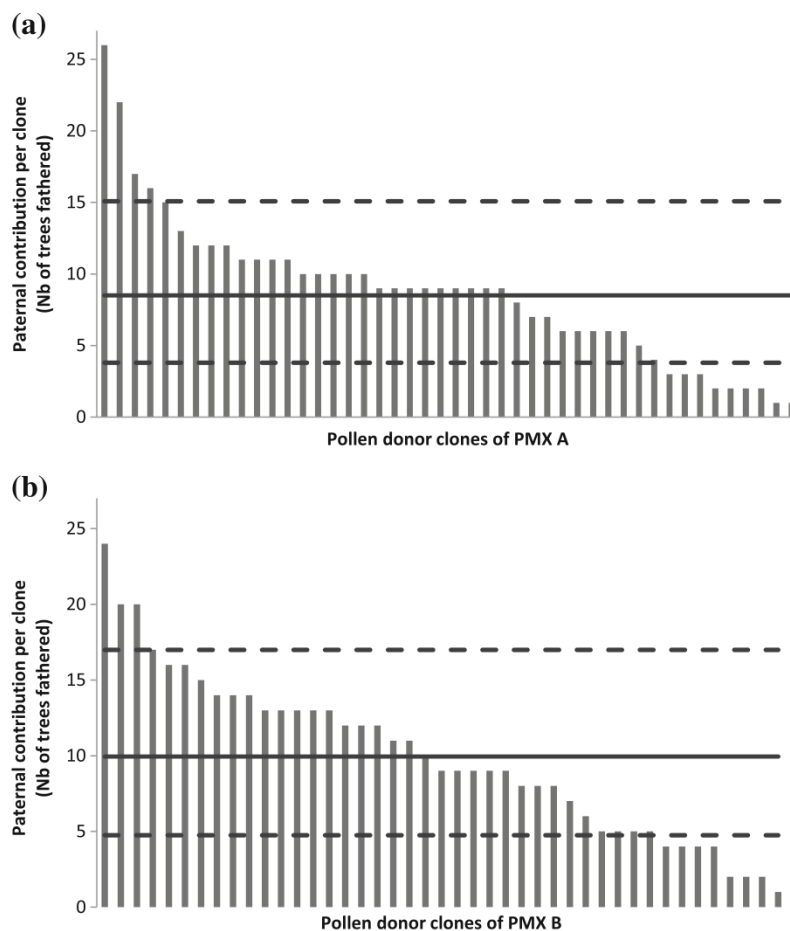
### Reproductive success of pollen parents

All the genotyped pollen donors were represented in the progeny pool (except the single pollen donor not genotyped belonging to PMX\_A), although their contributions were variable.

Deviation from expected equal male reproductive success across all females was tested using the chi-square test (only offspring sired by a pollen donor from the PMX were considered for testing the null hypothesis of equal male contribution). The expected frequencies were obtained by dividing the number of offspring correctly assigned by the number of pollen parents in each PMX. In Fig. 1, the paternal contribution of each pollen donor is represented for both PMX, and the expected equal contribution is indicated by a continuous dark line. In Subset\_A (Fig. 1a), paternal contribution varied from 1 to 26 (0.3 to 6.5 %), across all female parents, whereas the expected paternal equal contribution was 8.5 (2.1 %). In Subset\_B (Fig. 1b), paternal contribution varied from 1 to 24 (0.23 to 5.61 %), whereas the expected paternal equal contribution was 9.95 (2.33 %). The null hypothesis that each pollen donor contributes equally to seed production was thus rejected in Subset\_A ( $\chi^2 = 149$ ;  $p = 8.6e-13$ ) and Subset\_B ( $\chi^2 = 121$ ;  $p = 1.2e-09$ ). However, it should be noted that only few pollen donors were out of the confidence interval (at 95 % probability) indicated by dashed lines in Fig. 1. Four pollen donors pollinated more than expected in both subsets, while 9 and 8 in Subset\_A and Subset\_B, respectively, contributed less than expected. Thus, the majority of pollen donors fell within the expected range of variation.



**Fig. 1** Paternal contribution of the 47 and 43 pollen donors of PMX\_A (a) and PMX\_B (b). The average expected contribution (continuous line) and 95 % confidence intervals (dashed lines) are indicated. Each vertical bar represents a different pollen donor of the polymix



Moreover, the disequilibrium in paternal contributions did not generate an important reduction in genetic variability. Indeed, the ratio between the observed and the expected (if equal contribution) effective number of males ( $N_e m$ ) was 0.81 for PMX\_A and 0.84 for PMX\_B. Therefore, the observed differential reproductive success did not lead to an important reduction of the effective number of males ( $N_e m$ ).

#### Proportions of relatives and coancestry among offspring

A major output of this paternity recovery experiment was the reconstruction of the complete mating design. Among the 400 offspring of Subset\_A, 352 different crosses could be identified (Fig. 2a): 313, 34, and 5 crosses were represented by 1, 2, and up to 6 offspring, respectively. In Subset\_B (Fig. 2b), 378 different crosses were identified among the 428 analyzed offspring: 331, 44, and 3 crosses were represented by 1, 2, and 3 offspring, respectively. Thus, in both subsets, PMX families (size: 9–12 offspring) contained between 0 to 6 full-sibs: 32 families consisted solely of half-sibs, and only 8 families comprised more than 3 full-sibs. Moreover, the selfing rate was

very low, in spite of 4 and 10 pollen donors in PMX\_A and PMX\_B, respectively, which also corresponded to seed parents. Only two offspring, belonging to Subset\_A, were issued from a self-fertilization of the same genotype.

Thus, for most of the HS families studied, the estimated coancestry between progeny within a family did not greatly differ from the expected additive relationship coefficient  $2 = 0.25$ . Indeed, mean coancestry among offspring within maternal families with full pedigree information was 0.260 (ranging from 0.250 to 0.365 according to the families). The family with the highest average coancestry ( $2 = 0.365$ ) contained two progeny with an additive relationship coefficient equal to 1 because they were both issued from the selfing of the same female parent.

Moreover, because some G1 parents were related, some relationships among offspring between families were found (Fig. 3). For example, in the partial pedigree model, some individuals had one grandparent in common ( $2 = 0.0625$ ) and some others had two grandparents in common ( $2 = 0.125$ ). Paternity recovery brought more complexity: relatedness among parents increased the coancestry among offspring within and between families.

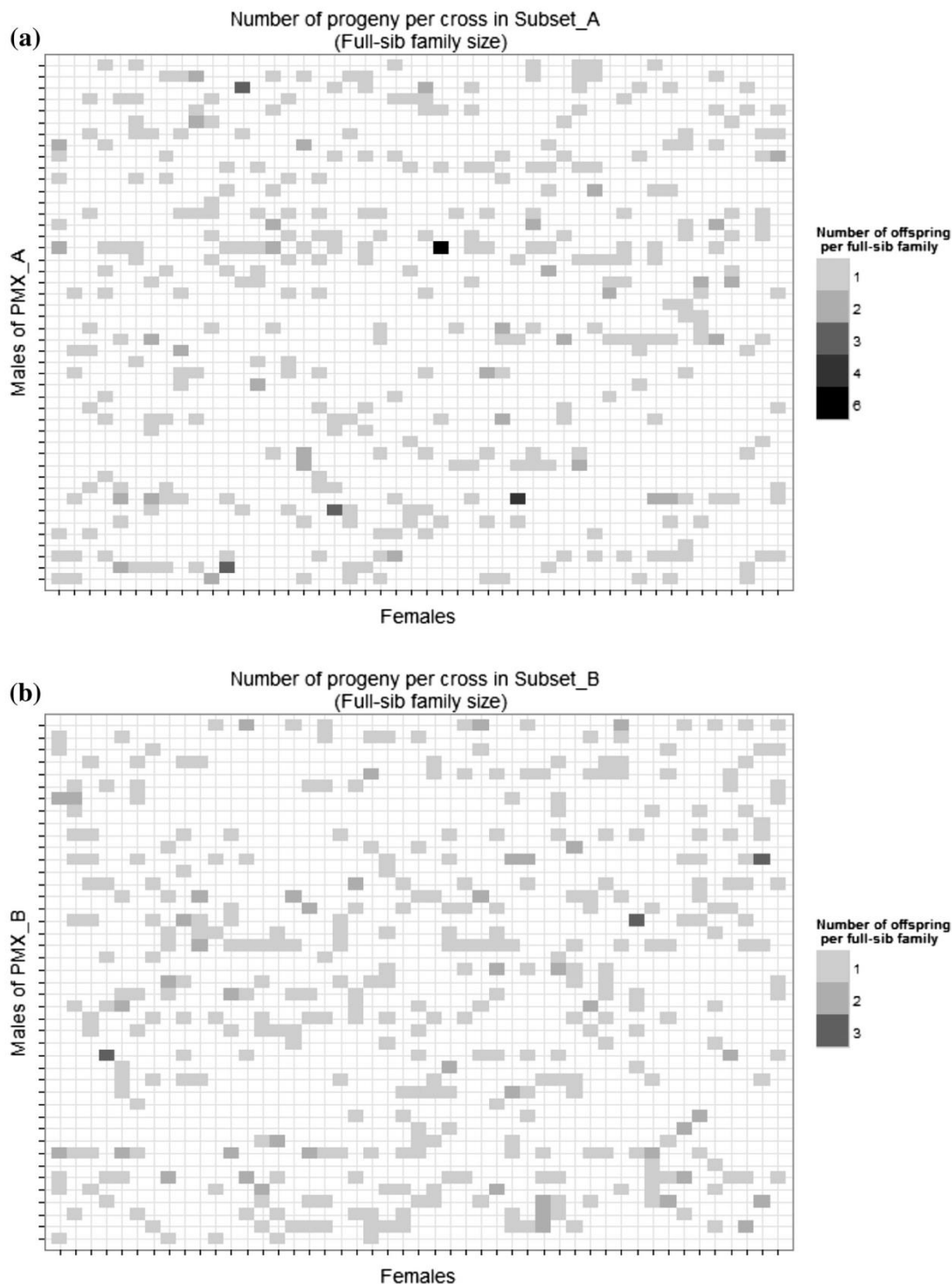
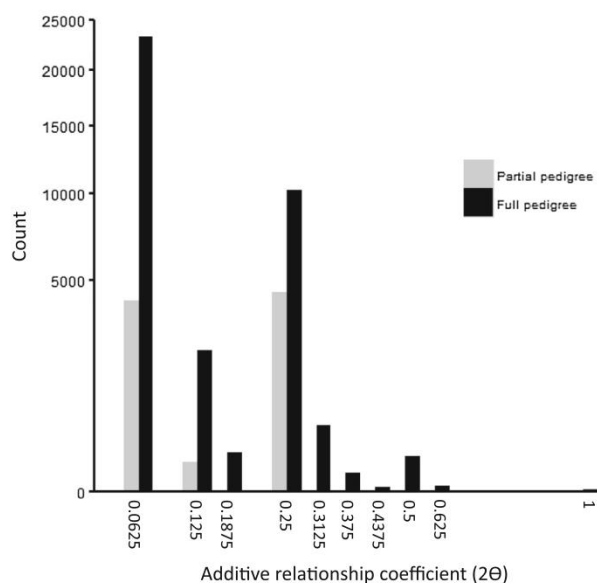


Fig. 2 Mating design revealed by pedigree recovery of the 400 and 428 crossings from Subset\_A (a) and Subset\_B (b), respectively

**Genetic parameters estimation**

Genetic parameters were estimated with a partial pedigree (PP) model and a full pedigree (FP) model for the 929 G2

genotyped trees with confirmed seed donors. Subset\_A and Subset\_B were thus considered together in these analyses. Additive variance and heritability estimates for girth at the age of 12 years and sweep at the age of 8 years for each model



**Fig. 3** Distribution of additive relationship coefficients ( $2\theta$  from A matrix of size 929) among genotyped G2 trees for partial and full pedigree models. The number of  $2\theta = 0$  was removed from this figure

are given in Table 2. Heritability estimates computed with the PP model were lower than heritability estimates computed with the FP model for both traits analyzed. The PP model tended to overestimate heritability, especially for sweep. However, the difference between heritability estimated with PP and FP models were not significant for both traits, as shown by overlapping confidence intervals at 95 % probability. The standard errors associated were slightly smaller when the FP model was used.

#### Estimated breeding values for parental population

Correlation between breeding values predicted using PP and FP models were high for both traits for G1 trees used as seed donors (Fig. 4). Pearson’s correlation coefficients were 0.97 for sweep and 0.96 for girth. These high correlations indicated that female breeding values were adequately predicted using the PP model for most of the individuals. Moreover, the high value of Spearman rank correlations between EBVs predicted by FP or PP models (0.97 for sweep and 0.96 for girth)

revealed that the ranking of seed donors was similar with both models, indicating that backward selection would be similar whatever the model (PP or FP). However, for a few seed donors, the estimation of breeding values varied substantially when the paternal information was added, and this could not be related to a high level of coancestry within the offspring (seed donors where the within-family average relatedness,  $2\theta$ , is greater than 0.27 are represented with closed circles in Fig. 4).

The accuracy of mean EBVs for G1 trees used as seed donors (female EBV) was similar for both models (Table 2). Thus, paternity recovery did not improve the accuracy of parental EBVs for seed donors. However, paternity recovery provided information for genetic evaluation of genotypes used as males, meaning that males could also be included in the backward selection pool. Nevertheless, mean accuracy for genotypes within the PMX ( $r=0.68$  for sweep;  $r=0.61$  for girth) was slightly lower than for seed donors ( $r=0.70$  for sweep;  $r=0.63$  for girth). Furthermore, EBV accuracy varied substantially between genotypes used as males: it ranged from 0.32 to 0.85 for sweep and from 0.27 to 0.80 for girth, depending on the number of progeny per male.

## Discussion

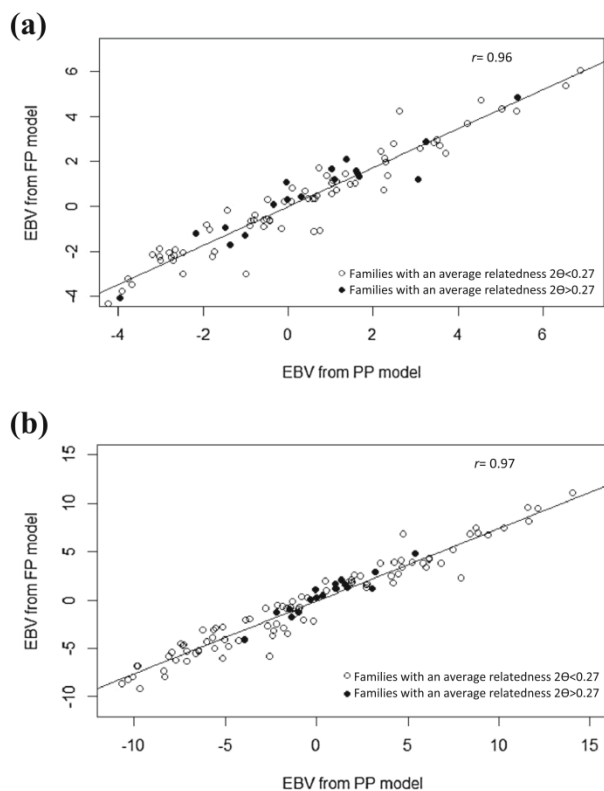
### Development of a cost-efficient genotyping technology for pedigree recovery in maritime pine

According to Glaubitz et al. (2003), “at least 32 independently segregating SNPs (each with a minor allele frequency of 0.2 or greater) are required to achieve an exclusion probability greater than 0.99” when one parent is known. In our study, paternity analysis was based on 42 SNPs or more (with MAF > 0.45). Thus, our set of markers had very strong discrimination power (probability of exclusion close to 1). This led us to consider a 99 % confidence level to determine the delta score used as a criterion for assignment of paternity. This confidence level was higher than the levels commonly used in most of other pedigree recovery studies (95 % as high or “strict” confidence and 80 % as good or “relaxed” confidence were often

**Table 2** Results from ASReml analyses for girth (age 12) and stem sweep (age 8)

Trait	Model	Mean	Additive var	SE	CVa	Heritability (SE)	$r_{(BV_s)}$
Girth	PP	53.37	15.90	6.10	0.07	0.25 (0.09)	0.63
	FP	53.37	13.54	3.98	0.07	0.21 (0.06)	0.63
Sweep	PP	19.02	65.57	17.37	0.43	0.47 (0.11)	0.75
	FP	19.02	43.29	10.41	0.35	0.31 (0.07)	0.71

Comparison of additive variance components (and corresponding standard errors SE), coefficient of variation (CVa), heritability estimates (and their associated standard errors), and accuracy of the female predicted breeding values ( $r$ ), according to the pedigree information model (partial pedigree (PP) and full pedigree (FP))



**Fig. 4** Correlation between female EBVs from partial pedigree (PP) and full pedigree (FP) models for girth (a) and stem sweep (b).  $r$  is the Pearson product moment correlation coefficient. Closed circles represent the 18 families with the highest level of within family coancestry (*i.e.* the 18 seed donors with a family-average relatedness ( $2\theta$ ) higher than 0.27)

used) (Doerksen and Herbinger 2008; Doerksen and Herbinger 2010; El-Kassaby et al. 2011; Funda et al. 2014; Grattapaglia et al. 2014; Hansen and McKinney 2010; Hansen and Nielsen 2010; Klapste et al. 2014; Lai et al. 2010).

Moreover, because the pollen donors were not related in PMX\_B and poorly related in PMX\_A (coancestry coefficient of 0.036 in PMX\_A and 0 in PMX\_B), discrimination power was thus enhanced. To conclude, the optimized set of 63 SNPs developed in this study for maritime pine provides a reliable genotyping tool for pedigree recovery, at least for the Landes provenance. The genotyping cost using Sequenom technology was less than €7 per individual. This set of markers is now used routinely for the French maritime pine breeding program: for example, to check pedigrees or tree identities or to evaluate pollution rates in seed orchards.

#### High rate of pollen contamination

We consider here that the pollen contamination rate is the proportion of progeny with no father identified within the PMX composition. In this study, 13.0 and 8.7 % of the progeny were fathered by outside pollen in Subset\_A and Subset\_B, respectively. However, this contamination rate

was overestimated in Subset\_A since one pollen donor belonging to PMX\_A was not genotyped (genotype not available). However, if we assign the average number of expected progeny (*i.e.* 8.5 offspring) to this pollen donor, the contamination rate of Subset\_A decreases only to 11.2 %.

The contamination rate varied greatly according to the family considered. It was higher than contamination rates usually reported in polycross trials in the literature: 4.1 % in red spruce (Doerksen and Herbinger 2008), 5 % in poplar (Wheeler et al. 2006), 13 % in radiata pine (Kumar et al. 2007), and no contamination in Japanese cedar, except in one family with a contamination rate of 22 % (Moriguchi et al. 2009).

Numerous hypotheses could explain this high rate of contamination. For the highly contaminated families (for example, contamination rate of 70 % for 1 family belonging to Subset\_A), it is likely that there was a problem with flower isolation. Other possibilities include the collection of pollens on parents with erroneous identities in the clonal archives. Indeed, mishandled pollen or mislabeled pollen donor identities were possible and some foreign pollens could have been introduced accidentally in the PMX, contributing to increase of contamination. Some studies have already reported mislabeling errors in clonal archives, seed orchards, and trials. In a detailed study using allozymes in conifer tree improvement programs in the USA, Adams et al. (1988) found that about 30 % of the controlled crosses in Douglas fir and loblolly pine had at least two progeny different from expectations. Mislabeled individuals during various field operations was the likely explanation for discrepancies of this sort. In a red spruce polycross trial, three types of pedigree errors were detected (Doerksen and Herbinger 2008): volunteers (“wild tree unconnected to the pedigree detected as individual progeny mismatching at more than two loci with their supposed mother”), mishandling, and foreign male pollination. Kumar and Richardson (2005) reported that about 5 % of the documented relationships (*e.g.*, full-sibs, parent-offspring, etc.) in the New Zealand radiata pine breeding population were identified as errors by molecular markers. In a black walnut progeny test, Zhao et al. (2013) showed that 20 % of the genotyped seedlings were not assigned to their theoretical families and that these misidentifications had a significant effect on the rank of families and potential selection of individuals. They recommended to genotype progeny soon after trial establishment or at least to genotype all selections to verify their identity and pedigree in order to avoid perpetuating an error into future generations and confounding breeding plans. Finally, it is important to note that in this study, the G2 individuals analyzed constituted the third generation of the French maritime pine breeding program. So, errors during various field operations as pollen harvesting, grafting in clonal archives, or controlled crossing were accumulated during the three selection cycles leading to different types of mislabeling. This mislabeling hypothesis was

strengthened by the 41 genotyped offspring which did not conform to their nominal seed donor genotypes.

#### Low departure from equal male reproductive success

Both subsets came from two different PMX, but the distribution of paternal contributions was similar in both cases (Fig. 1). Even though the equal paternal contribution hypothesis was rejected in both subsets, a homogenous contribution close to the expected mean was observed, and only a few males contributed more or less than expected. This result was in accordance with the majority of pedigree recovery studies that had demonstrated that paternal contributions in controlled polycross trials were unequal (Moriguchi et al. 2009). Wheeler et al. (2006) studied male reproductive success in a 16-parent pollen polymix for *Populus*: 50 % of the pollen donors contributed to 83 % of the progeny, and the contribution of each father varied from 1 to 13 %. Doerksen and Herbinger (2008) reported that in a polycross red spruce progeny test, the 21 pollen donors constituting the polymix had unequal reproductive success: while the expected paternal equal contribution was 4.8 %, each pollen donor contributed from 1.4 to 11 % to the progeny. One father contributed more than twice as much as had been expected; three fathers contributed less than twice as much as had been expected. Thus, our results agree with these studies: in Subset\_A and Subset\_B respectively, 2 and 1 pollen donors contributed more than expected and 10 and 8 pollen donors contributed less than expected.

The causes of unequal pollen contribution have been studied in a large number of species. Aronen et al. (2002), Snow and Spira (1996), and Pasonen et al. (1999) reported that pollen-tube growth rate affects paternal contribution in *P. abies*, *Hibiscus moscheutos*, and *B. pendula*, respectively. Moreover, Parantainen and Pasonen (2004) found pollen–pollen interactions in artificial crossing experiments using pollen mixtures in *Pinus sylvestris*. Nakamura and Wheeler (1992) reported that genetic incompatibility between male and female gametophytes affects paternal contributions in *P. menziesii*. In addition, Moriguchi et al. (2009) found that the contribution of each pollen donor in a polycross was related to their germination rate, but Wheeler et al. (2006) reported that paternal success appeared to be unrelated to pollen vigor.

Moreover, it is known that the number of males constituting the pollen mix has an impact on the differential reproductive success (DRS) (Kumar et al. 2007; Moriguchi et al. 2009). Therefore, using a large number of pollen donors is a general recommendation to decrease variations in the level of paternal contribution. In radiata pine, for example, Kumar et al. (2007) recommended using at least 15 males. Finally, the DRS could also be explained by genetic incompatibility (pre- and/or post-zygotic reproductive barriers). But in our study, the sampling did not contain enough progeny per

family compared to the number of potential pollen parents to solve this question.

It is also important to note that since only healthy and vigorous seedlings are typically planted in progeny trials and generally not all seedlings survive by the selection age, the relative contributions of pollen parents could deviate from those observed at the sound-seeds stage (Kumar et al. 2007).

Pedigree reconstruction showed that a large number of crosses were actually realized with the polycross mating design, even if few individuals were available for each cross. The 828 progeny analyzed represented 730 different crosses, and 88 % of them were realized once (*i.e.* a unique progeny/cross). Despite the DRS of pollen parents, the HS family assumption did not deviate from the null hypothesis and only a few families contained several full-sibs.

Finally, the selfing rate was very low: only two progeny were issued from self-mating, while there was a significant possibility of obtaining selfed offspring (4 and 10 pollen donors in PMX\_A and PMX\_B, respectively, also appeared in the seed donor list). This was not surprising because self-fertilization causes inbreeding depression in conifers, associated with high seedling mortality at an early stage, and reduced vigor in plants (Chancerel et al. 2013). In maritime pine, most lethal or sub-lethal alleles are probably eliminated during seed formation and germination as well as during the first growing season so that the proportion of selfed offspring is expected to be very low in mature populations (Gonzalez-Martinez et al. 2003). Durel et al. (1996), however, noticed that maritime pine survival rates after a first growing season in the nursery were similar independently of the level of inbreeding and also observed that there were large differences of inbreeding depression among inbred families for the same level of inbreeding. In our study, the offspring genotyped were 16 years old at the time of pedigree recovery, so several selection steps could have occurred before our analysis (seed development, germination, plantation, etc.), possibly resulting in a lower selfing rate than that at seed stage. Noteworthy, only 10 offspring per family were genotyped in our study, so the number of offspring was lower than the number of potential pollen parents. Thus, our sampling was not really appropriate to study the selfing rate and it is likely that the HS family assumption would have deviated more significantly from the null hypothesis if more offspring per family had been genotyped.

#### Low bias in heritability and EBV due to unknown paternity

In most forest tree breeding programs, genetic parameter estimation relies on open-pollinated or polycross testing schemes. To evaluate the breeding value of selected parents, one can proceed by collecting seeds from desirable parents and establishing HS progeny trials, where a particular parent's breeding

value is calculated as the deviation of its progeny mean from the population mean weighted by heritability. These estimations assume true HS families and equal contribution of male parents. Here, we studied the bias in genetic parameters estimation by comparison of heritability estimates calculated with and without paternity information and by calculating the correlation between female EBVs estimated by one model or the other.

Despite the unequal contribution of pollen parents, the average relatedness within families was 0.26 (slightly superior to the additive relationship coefficient of 0.25 expected for half-sibs) and the estimation of heritability was not significantly overestimated for girth and sweep. These trends were already reported in the literature. For example, Doerksen and Herbinger (2010) concluded that in a red spruce PMX trial, the heritability estimate increased slightly when adding pedigree information, mostly due to the capture of genetic variations between males (and not due to the correction of within-family genetic correlation coefficient). Other studies found either no significant bias in heritability estimation due to the inclusion of full-sibs information (Hansen and Nielsen 2010), or an overestimation of heritability estimates (El-Kassaby et al. 2011; Gauzere et al. 2013; Hansen and Nielsen 2010; Korecky et al. 2013) when paternity information was unknown. It should be stated that the sampling effect, the population size, the trait considered, and the number of parents involved, all have important effects on heritability estimates.

Moreover, we showed that breeding values estimated with and without pollen parent identity were highly correlated. We also noticed that the presence of full-sibs within families did not consistently lead to a biased estimation of breeding values. So ranking obtained from polycross designs (without paternity recovery), such as in this study, would be satisfactory to implement an efficient backward selection, provided that the number of pollen parents constituting the PMX is large. Doerksen and Herbinger (2010) reported that correlations between female EBVs estimated through partial and full pedigree models were close to one in a red spruce PMX design. Other studies with open-pollinated families (Hallingback and Jansson 2013; Hansen and Nielsen 2010; Klapste et al. 2014; Korecky et al. 2013) concluded that the ranking of the females based on their EBVs estimated from partial and full pedigree models was very similar as well.

In many tree breeding programs, the use of polycross designs is restricted to the estimation of female EBVs only because the lack of information on the male parentage strongly restrains any advanced-generation selection. Lambeth et al. (2001) proposed the use of molecular markers to identify the male parents of forward selections, allowing an enhanced application of polycross designs. They presented practical scenarios for using molecular markers to identify male parentage in polycross designs. Successful implementation of polymix

breeding and testing system requires an accurate pedigree reconstruction, low differential in male reproductive success, low pollen contamination rate (since selection is performed only when a complete pedigree is known), and a low selfing rate. This study demonstrates that our polycross progeny trial complied with most of these requirements. Based on these results, implementing a polymix breeding strategy in the maritime pine breeding program will be investigated in the near future.

To conclude, we developed an affordable SNP marker technology providing accurate pedigree recovery for maritime pine. We showed that the pollen contamination rate was rather high (11 %) and that the differential reproductive success of pollen parents used in polymixes was significant due to a few pollen donors contributing more or less than expected. We also demonstrated that heritability and breeding value estimates in this polycross progeny trial are not biased when the pollen donor identity is unknown. Finally, despite recovering paternal identities, this did not result in a significant improvement for backward selection; the new pedigree information enabled forward selection of offspring in the maritime pine breeding program and opened new perspectives for future breeding strategies.

**Acknowledgments** This study was funded by INRA (EFPA division *projet innovant*), the European-Union (ProCoGen project: n° 289841), and *Conseil Régional d'Aquitaine* (IMAF project co-funded by FCBA: n° 120009468–052). Marjorie Vidal received a CIFRE PhD fellowship (public/private research partnerships between FCBA and the French Ministry of Higher Education and Research). This study would not have been possible without the support of the maritime Pine Breeding Cooperative (*GIS Pin Maritime du Futur*). We gratefully acknowledge all its members. The authors also thank the INRA Experimental Unit (UE0570) for field measurements, Sophie Gerber (INRA) for helpful discussions on paternity recovery, Adline Delcamp (INRA) for the genotyping stage, Jean-Mathieu De Boissesson (FCBA) for needle sampling, Maël Ruby for his preliminary study on paternity recovery, and Fikret Isik (NCSU, USA) for his advice in shaping the content of the paper and his help in ASReml analysis.

The genotyping was performed at the Genomic Facility of Bordeaux (grants from the *Conseil Régional d'Aquitaine*: n°20030304002FA and 20040305003FA, the European Union: FEDER n°2003227 and ANR: n°ANR-10-EQPX-16 Xyloforest).

**Data archiving statement** All SNPs used in this study were deposited in dbSNP (NCBI database), available on <http://www.ncbi.nlm.nih.gov/>.

The supplemental data “Supplemental\_data\_SNP\_description.txt” contains the dbSNP accession ID for each SNP (#ss) and some useful information (SNP ID, set identity, contig name, linkage group LG, position on the LG, SNP type, and sequence surrounding the SNP).

The genotypes of G2 trees and their potential parents (G1 trees) are available in the supplemental data file “Subset\_A\_genotypes\_56SNPs.csv” for Subset A and “Subset\_B\_genotypes\_63SNPs.csv” for subset B. The first column contains the individual IDs, and the first line contains the SNP IDs. Blank cells are missing data.

## References

- Adams WT, Neale DB, Loopstra CA (1988) Verifying controlled crosses in conifer tree-improvement programs. *Silvae Genet* 37:147–152
- Aronen T, Nikkanen T, Harju A, Tiimonen H, Haggman H (2002) Pollen competition and seed-siring success in *Picea abies*. *Theor Appl Genet* 104:638–642. doi:10.1007/s00122-001-0789-9
- Burdon RD, Shelbourne CJA (1971) Breeding populations for recurrent selection: conflicts and possible solutions. *N Z J For Sci* 1:174
- Canales J, Bautista R, Label P, Gomez-Maldonado J, Lesur I, Fernandez-Pozo N, Rueda-Lopez M, Guerrero-Fernandez D, Castro-Rodriguez V, Benzekri H, Canas RA, Guevara MA, Rodrigues A, Seoane P, Teyssier C, Morel A, Ehrenmann F, Le Provost G, Lalanne C, Noirot C, Klopp C, Reymond I, Garcia-Gutierrez A, Trontin JF, Lelu-Walter MA, Miguel C, Cervera MT, Canton FR, Plomion C, Harvenget L, Avila C, Gonzalo Claros M, Canovas FM (2014) De novo assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. *Plant Biotechnol J* 12:286–299. doi:10.1111/pbi.12136
- Chancerel E, Lamy JB, Lesur I, Noirot C, Klopp C, Ehrenmann F, Boury C, Le Provost G, Label P, Lalanne C, Leger V, Salin F, Gion JM, Plomion C (2013) High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biol* 11:19. doi:10.1186/1741-7007-11-50
- Doerksen TK, Herbinger CM (2008) Male reproductive success and pedigree error in red spruce open-pollinated and polycross mating systems. *Can J For Res-Rev Can Rech For* 38:1742–1749. doi:10.1139/x08-025
- Doerksen TK, Herbinger CM (2010) Impact of reconstructed pedigrees on progeny—test breeding values in red spruce. *Tree Genet Genomes* 6:591–600. doi:10.1007/s11295-010-0274-1
- Durel CE, Bertin P, Kremer A (1996) Relationship between inbreeding depression and inbreeding coefficient in maritime pine (*Pinus pinaster*). *Theor Appl Genet* 92:347–356
- El-Kassaby YA, Cappa AP, Liewlaksaneeyanawin C, Klapste J, Lstiburek M (2011) Breeding without breeding: is a complete pedigree necessary for efficient breeding? *PLoS One* 6:11. doi:10.1371/journal.pone.0025737
- Funda T, Liewlaksaneeyanawin C, El-Kassaby YA (2014) Determination of paternal and maternal parentage in lodgepole pine seed: full versus partial pedigree reconstruction. *Can J For Res-Rev Can Rech For* 44:1122–1127. doi:10.1139/cjfr-2014-0145
- Gaspar MJ, de-Lucas A, Alia R, Paiva JAP, Hidalgo E, Louzada J, Almeida H, Gonzalez-Martinez SC (2009) Use of molecular markers for estimating breeding parameters: a case study in a *Pinus pinaster* Ait. progeny trial. *Tree Genet Genomes* 5:609–616. doi:10.1007/s11295-009-0213-1
- Gauzere J, Oddou-Muratorio S, Pichot C, Lefevre F, Klein E (2013) Biases in quantitative genetic analyses using open-pollinated progeny tests from natural tree populations. *Acta Bot Gall* 160:225–236. doi:10.1080/12538078.2013.822827
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK
- Glaubitz JC, Rhodes OE, Dewoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol Ecol* 12:1039–1047. doi:10.1046/j.1365-294X.2003.01790.x
- Gonzalez-Martinez SC, Gerber S, Cervera MT, Martinez-Zapater JM, Alia R, Gil L (2003) Selfing and sibship structure in a two-cohort stand of maritime pine (*Pinus pinaster* Ait.) using nuclear SSR markers. *Ann For Sci* 60:115–121. doi:10.1051/forest:2003003
- Grattapaglia D, Diener PSD, dos Santos GA (2014) Performance of microsatellites for parentage assignment following mass controlled pollination in a clonal seed orchard of loblolly pine (*Pinus taeda* L.). *Tree Genet Genomes* 10:1631–1643. doi:10.1007/s11295-014-0784-3
- Hallingback HR, Jansson G (2013) Genetic information from progeny trials: a comparison between progenies generated by open pollination and by controlled crosses. *Tree Genet Genomes* 9:731–740. doi:10.1007/s11295-012-0588-2
- Hansen OK, McKinney LV (2010) Establishment of a quasi-field trial in *Abies nordmanniana*—test of a new approach to forest tree breeding. *Tree Genet Genomes* 6:345–355. doi:10.1007/s11295-009-0253-6
- Hansen OK, Nielsen UB (2010) Microsatellites used to establish full pedigree in a half-sib trial and correlation between number of male strobili and paternal success. *Ann For Sci* 67:10. doi:10.1051/forest/2010028
- Illy G (1966) Recherches sur l'amélioration génétique du pin maritime. *Ann Sci Forest* 23:765–948. doi:10.1051/forest/19660401
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16:1099–1106. doi:10.1111/j.1365-294X.2007.03089.x
- Kimura M, Crow JF (1963) The measurement of effective population number. *Evolution* 17:279–288
- Klapste J, Lstiburek M, El-Kassaby YA (2014) Estimates of genetic parameters and breeding values from western larch open-pollinated families using marker-based relationship. *Tree Genet Genomes* 10:241–249. doi:10.1007/s11295-013-0673-1
- Korecky J, Klapste J, Lstiburek M, Kobliha J, Nelson CD, El-Kassaby YA (2013) Comparison of genetic parameters from marker-based relationship, sibship, and combined models in Scots pine multi-site open-pollinated tests. *Tree Genet Genomes* 9:1227–1235. doi:10.1007/s11295-013-0630-z
- Kumar S, Gerber S, Richardson TE, Gea L (2007) Testing for unequal paternal contributions using nuclear and chloroplast SSR markers in polycross families of radiata pine. *Tree Genet Genomes* 3:207–214. doi:10.1007/s11295-006-0056-y
- Kumar S, Richardson TE (2005) Inferring relatedness and heritability using molecular markers in radiata pine. *Mol Breed* 15:55–64. doi:10.1007/s11032-004-2059-4
- Lai BS, Funda T, Liewlaksaneeyanawin C, Klapste J, Van Niejenhuis A, Cook C, Stoehr MU, Woods J, El-Kassaby YA (2010) Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction. *Ann For Sci* 67:8. doi:10.1051/forest/2010044
- Lambeth C, Lee BC, O'Malley D, Wheeler N (2001) Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet* 103:930–943. doi:10.1007/s001220100627
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc., Sunderland, MA, USA. **980 p**
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108:285–291. doi:10.1038/hdy.2011.73
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7:639–655. doi:10.1046/j.1365-294x.1998.00374.x
- Moriguchi Y, Ishiduka D, Kaneko T, Ito S, Taira H, Tsumura Y (2009) The contribution of pollen germination rates to uneven paternity among polycrosses of *Cryptomeria japonica*. *Silvae Genet* 58:139–144
- Nakamura RR, Wheeler NC (1992) Pollen competition and paternal success in Douglas-fir. *Evolution* 46:846–851. doi:10.2307/2409655
- Parantainen A, Pasonen HL (2004) Pollen-pollen interactions in *Pinus sylvestris*. *Scand J Forest Res* 19:199–205. doi:10.1080/02827580410029336
- Pasonen HL, Pulkkinen P, Kapyla M, Blom A (1999) Pollen-tube growth rate and seed-siring success among *Betula pendula* clones. *New Phytol* 143:243–251. doi:10.1046/j.1469-8137.1999.00451.x
- Plomion C, Chancerel E, Endelman J, Lamy JB, Mandrou E, Lesur I, Ehrenmann F, Isik F, Bink M, van Heerwaarden J, Bouffier L (2014)

#### 4. Reconstitution de pedigree dans un test de descendance polycross et conséquences pour la stratégie de sélection backward

- Genome-wide distribution of genetic diversity and linkage disequilibrium in a mass-selected population of maritime pine. *BMC Genomics* 15:16. doi:10.1186/1471-2164-15-171
- RCoreTeam (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Snow AA, Spira TP (1996) Pollen-tube competition and male fitness in *Hibiscus moscheutos*. *Evolution* 50:1866–1870. doi:10.2307/2410744
- Squillace AE (1974) Average genetic correlations among offspring from open-pollinated forest trees. *Silvae Genetica* 23:149–156
- Wheeler N, Payne P, Hipkins V, Saich R, Kenny S, Tuskan G (2006) Polymix breeding with paternity analysis in *Populus*: a test for differential reproductive success (DRS) among pollen donors. *Tree Genet Genomes* 2:53–60. doi:10.1007/s11295-005-0033-x
- White T (1987) A conceptual framework for tree improvement programs. *New For* 1:325–342. doi:10.1007/BF00031742
- Wiseloge AE, Vanbuijtenen JP (1988) Probability of equal mating in polymix pollinations of loblolly pine (*Pinus taeda* L.). *Silvae Genet* 37:184–187
- Zhao P, Zhang SX, Woeste K (2013) Genotypic data changes family rank for growth and quality traits in a black walnut (*Juglans nigra* L.) progeny test. *New For* 44:357–368. doi:10.1007/s11056-012-9343-7
- Zobel BJ, Talbert JT (1984). *Applied forest tree improvement* (p. 528). New York: Wiley.



### 4.3 Conclusion de la partie 4

L'**objectif I** consistait à développer un jeu de marqueurs moléculaires permettant de reconstituer les pedigrees dans la population d'amélioration du pin maritime de provenance landaise. Deux multiplex de 40 SNPs chacun ont été définis pour génotyper à l'aide de la technologie Sequenom® iPLEX® Gold les individus à assigner ainsi que leurs parents potentiels. Au final, 63 SNPs ont été retenus pour les analyses de parenté (réalisées à l'aide du logiciel Cervus). Ce jeu de marqueurs est d'ailleurs en cours d'utilisation pour d'autres études (*cf. partie 6.1*) au sein du programme d'amélioration.

L'**objectif II** consistait à étudier l'impact de la connaissance d'un pedigree complet sur une possible optimisation de la stratégie de sélection *backward*. On a montré que les familles étudiées étaient majoritairement composées de demi-frères, et que la majorité des pères constituant le polymix contribuaient de façon équivalente à la descendance (même si certains pères contribuent plus ou moins que ce qui est attendu). Ainsi, les hypothèses sur lesquelles repose la sélection *backward* sont bien respectées dans le test de descendance polycross étudié. Ensuite, on a estimé les paramètres génétiques et les valeurs génétiques des mères G1 évaluées, avec et sans l'information du pedigree complet. On a montré que le classement des mères G1 (basé sur les valeurs génétiques) est peu modifié lorsqu'on apporte l'information d'un pedigree complet. Ainsi, l'évaluation génétique de la sélection *backward* classique n'est pas biaisée du fait d'un pedigree partiel. Toutefois, la reconstitution de pedigree permet d'améliorer la précision de l'évaluation génétique et de corriger les erreurs d'identité et de pedigree.

De plus, la reconstitution de pedigree dans un test de descendance polycross ouvre des perspectives intéressantes pour la sélection *forward* : les descendants G2 génotypés avec un pedigree complet pourraient être sélectionnés pour constituer la population d'amélioration de la génération suivante et/ou pour établir une population de production de graines améliorées (vergers à graines). Ces perspectives sont évaluées dans la *partie 5*.

## 5. Evaluation d'une stratégie de sélection *forward* de type « *polymix breeding* »

Cette *partie 5* traite de l'*objectif III* de la thèse, qui consistait à tester la faisabilité d'une sélection forward au sein d'un test de descendance polycross (dispositif étudié préalablement dans la *partie 4*).

Cette partie a fait l'objet d'un article intitulé « *Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction* », soumis à *Annals of Forest Science* en janvier 2016.

### 5.1 Résumé de l'article

Cet article a pour objectif d'étudier la faisabilité d'une sélection *forward* au sein d'un test de descendance polycross. Pour cela, les descendants d'un test de descendance polycross ont été classés sur leurs valeurs génétiques (estimées avec l'information d'un pedigree partiel) pour le volume et la rectitude. Les meilleurs individus ont été présélectionnés en imposant ou non une restriction sur l'apparentement lors des présélections (*échantillonnage B*). Ils ont ensuite été génotypés avec 63 marqueurs SNPs (marqueurs développés dans la *partie 4*) afin de reconstituer leur pedigree. Les individus génotypés et présentant un pedigree complet étaient candidats à la sélection finale pour l'établissement d'une population de production (sélection en tant que géniteur en verger à graines). Ces individus candidats à la sélection ont été classés selon leurs valeurs génétiques estimées avec l'information du pedigree complet. Puis, différentes stratégies de sélection (*forward*, *backward* et mixte<sup>9</sup>) ont été testées en établissant par simulation des vergers à graines. Ces différentes stratégies ont été comparées en termes de gain génétique à niveau de diversité équivalent.

Ce travail montre que la sélection *forward* est envisageable pour valoriser les tests polycross existants. Le pedigree complet de la majorité des descendants présélectionnés et génotypés a été retrouvé. Les vergers à graines issus de leur sélection produisent une variété présentant des gains génétiques espérés légèrement supérieurs à ceux qu'on obtiendrait en réalisant une sélection *backward* classique. Le fait que la sélection *forward* n'offre pas un gain génétique vraiment supérieur à la sélection *backward* est probablement dû au fait que le test polycross étudié n'a pas été optimisé, notamment en terme de composition du polymix. De plus, les résultats montrent qu'il n'est pas nécessaire de mettre une contrainte sur l'apparentement lors de la présélection des individus (à génotyper), du moment que le nombre de candidats présélectionnés est assez grand pour garantir la diversité au sein du verger à graines. Par ailleurs, les gains génétiques obtenus sont similaires quel que soit le type de présélection choisi (avec ou sans contrainte sur l'apparentement).

---

<sup>9</sup> Mixte : dans cette stratégie de sélection, les individus sélectionnés sont issus à la fois de la sélection classique *backward* et de la sélection *forward* (on mélange des individus G1 et G2).

## 5. *Evaluation d'une stratégie de sélection forward de type polymix breeding*

Cette étude ouvre de nouvelles perspectives vers la mise en place d'une stratégie de sélection *forward* au sein du programme d'amélioration du pin maritime en Aquitaine, ce qui pourrait permettre d'accélérer les cycles de sélection, de simplifier la phase des croisements contrôlés et de maximiser le nombre de croisements réalisés.

### **5.2 Article 2**

## Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction.

Marjorie VIDAL<sup>1,2,3</sup>, Christophe PLOMION<sup>1,2</sup>, Annie RAFFIN<sup>1,2</sup>, Luc HARVENGT<sup>3</sup>, Laurent BOUFFIER<sup>1,2\*</sup>

<sup>1</sup> INRA, BIOGECO, UMR 1202, F-33610 Cestas, France;

<sup>2</sup> Univ. Bordeaux, BIOGECO, UMR 1202, F-33400 Talence, France;

<sup>3</sup> FCBA, Biotechnology and Advanced Silviculture Dept., Genetics & Biotechnology team, F-33610 Cestas, France

\* Corresponding author ([laurent.bouffier@pierroton.inra.fr](mailto:laurent.bouffier@pierroton.inra.fr))

### Key message

Molecular markers were used for paternity recovery in a polycross progeny trial, facilitating forward selection. Different breeding strategies for seed orchard establishment were evaluated by comparing genetic gains and diversity. This work opens up new perspectives in maritime pine breeding.

### Abstract

Polycross mating designs are widely used in forest tree breeding, to evaluate parental breeding values for backward selection. Alternatively, polycross progeny trials may be used to select the best trees on the basis of individual breeding values and molecular pedigree analysis. Such a forward selection strategy was tested, to provide proof-of-concept, in a maritime pine polycross trial. Progeny with higher breeding values for growth and stem straightness were first preselected with or without relatedness constraints. Then, preselected trees were genotyped for 63 single-nucleotide polymorphisms, to recover their pedigree. Finally, the preselected trees were ranked on the basis of their breeding values, estimated from the recovered full pedigree, and the best candidates were selected with three different strategies (forward, backward, mixed) and three levels of coancestry constraints (none, status number  $N_s=10$ ,  $N_s=20$ ), to establish a virtual clonal seed orchard. Genetic gains were estimated, for comparison of the various options. Complete pedigrees were recovered for most of the preselected trees. There was no major difference in expected genetic gains between the two preselection strategies. Genetic gains were slightly higher for forward selection than for classical backward selection. This seminal study opens up new perspectives for using forward selection within the French maritime pine breeding program.

**Keywords:** Breeding strategy, polymix breeding, paternity recovery, SNP markers, *Pinus pinaster* Ait.

## Introduction

Progeny testing for parental ranking is widely used in forest tree breeding (Zobel and Talbert 1984). In the French maritime pine breeding program, multisite polycross progeny trials have been established in the last 20 years, for assessing breeding values and ranking female parents for backward selection. The ranking of parents according to the performance of their progeny is particularly important for traits with a low heritability (Falconer and Mackay 1996), such as many traits of interest in forest trees (Burdon and Kumar 2004; Cornelius 1994; Pâques 2013). In such polycross trials, the progeny phenotyped to evaluate parental breeding values is not used to generate the next generation in the breeding population (which is actually selected from the progeny of biparental crosses) or included in production populations (commercial seed orchards). The lack of information about the male parent greatly limits the selection of advanced generation. However, progress in molecular genetics, such as the development of highly informative and cost-effective DNA markers, has made new approaches possible in tree breeding. One such new approach, pedigree reconstruction, makes it possible to reconstruct genealogical relationships between individuals, providing opportunities for the development of new breeding strategies. For example, controlled crosses can be replaced by pedigree recovery in open pollinated populations for the estimation of genetic parameters and the prediction of breeding values, in a strategy known as "Breeding without breeding" (El-Kassaby et al. 2011; El-Kassaby and Lstiburek 2009; Lstiburek et al. 2011). This approach can be used to initiate a tree improvement program without the need for the initial cycle of breeding and testing (Lstiburek et al. 2015). Another approach, "polymix breeding with parental analysis (PMX/WPA)" was developed by Lambeth et al. (2001) and combines controlled crosses and pedigree recovery. These authors proposed the use of molecular markers to identify the parents of potential selection candidates in polycross mating designs, for the evaluation of breeding values and the selection of progeny for the next generation in the breeding program. Three different scenarios are presented, depending on the progeny set genotyped: i) partial population paternity analysis (pedigree analysis only for the best progeny of the polycross trial); ii) full population paternity analysis (pedigree analysis for all progeny); iii) full-population parental analysis (identities of both female and male parents recovered by molecular marker analysis as the identities of the mothers are not recorded in this scenario, to decrease logistical costs). The authors claimed that PMX/WPA was a "viable alternative to full-sib breeding and testing system". Their approach presents several advantages. Polymix crosses are more cost-efficient than full-sib crosses for a given number of parents. They lead to a larger number of recombination events for fewer crosses and breeding values are more reliably estimated than for other methods, because each individual is crossed with a larger number of parents. In polycross progeny trials, the genetic gain from forward selection with PMX/WPA should be greater and it should be possible to deploy this gain more rapidly than that obtained with classical backward selection. Moreover, mislabeled clones can be eliminated in the genotyping phase, potentially increasing selection efficiency.

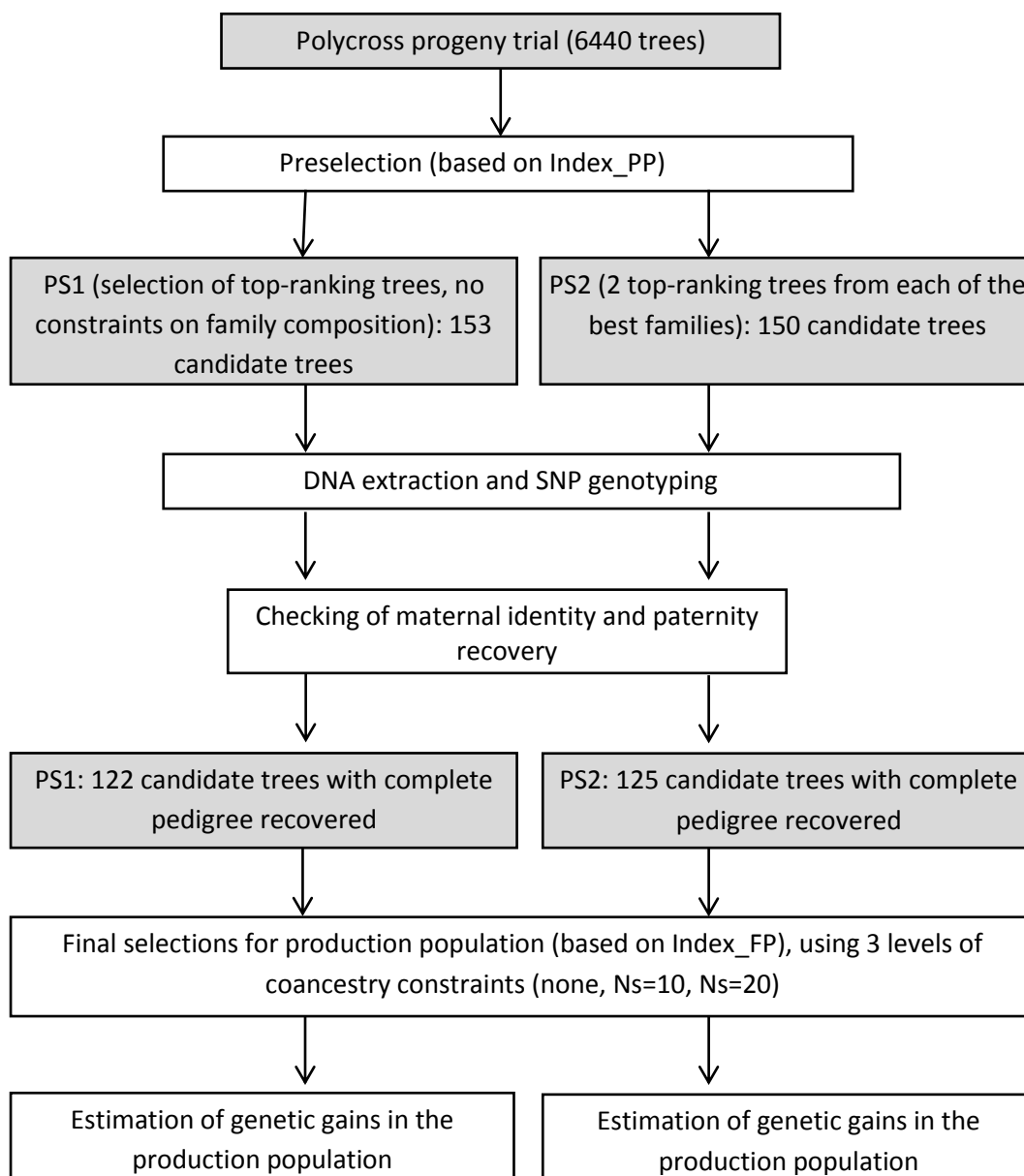
We investigated the feasibility of using a forward selection strategy in a maritime pine polycross trial associated with an analysis of the parentage of the progeny. Such forward selection may have two goals: recruitment of the best genotypes for the next generation of the breeding population and the formation of a production population, such as a clonal seed orchard (CSO). This study focuses on selection for the constitution of a CSO.

The successive stages of forward selection strategies studied here are presented in Fig. 1. Candidate trees were preselected in a polycross trial, using two different preselection options without (PS1) and with (PS2) constraints on relatedness. The preselected trees were then genotyped for single-nucleotide polymorphism (SNP) markers.

### 5. Evaluation d'une stratégie de sélection forward de type polymix breeding

Once their complete pedigrees were recovered, the best individuals were selected for the CSO, using three levels of coancestry constraints (none, status number  $N_s=10$ ,  $N_s=20$ ). Forward selection strategies were then compared, in terms of possible genetic gains, with backward and mixed (*i.e.* a combination of both) selection strategies, using the same three levels of coancestry constraints. Finally, ways of optimizing the preselection and final selection options for the establishment of a CSO were considered in the framework of the French maritime pine breeding program.

*Fig. 1 Main steps in forward selection (with two preselection options) in the maritime pine polycross trial.*



## Materials and Methods

### **Plant material and mating design**

Six polycross progeny trials for the maritime pine breeding program were established from 1994 to 2002 in South-Western France, for the prediction of second-generation (G1) parental breeding values. G1 trees were the selected progeny from G0 trees (plus trees), whereas G0 trees were mass-selected from the Landes provenance and constitute the base population of the breeding program (Illy 1966).

In total, 960 G1 trees were evaluated (as seed donors) within these six polycross progeny trials, each of which took place on three sites. This study focuses on one of these trial sites, established in 1996 (at 44°42'32"N / 0°46'8"W) for the evaluation of 166 G1 trees as seed parents. Two different pollen mixes were used: 98 seed parents pollinated with one polymix (47 G1 pollen donors,  $N_s=19$ ) and 76 seed parents pollinated with the other polymix (43 G1 pollen donors,  $N_s=43$ ). There were no pollen donors common to both polymixes. Eight of the G1 seed parents were pollinated with both polymixes, but the families resulting from identical seed parents and different PMX were considered to be different. The progeny trial thus consisted of 174 half-sib families plus five checklots, planted in a randomized block design, corresponding to a total of 6,440 trees (35 complete blocks with one tree plot per family and two trees per checklot in each block).

Progeny (G2 trees) of both polycrosses were phenotyped for selection criteria. Tree girth at breast height (GBH) and tree height (HT) were measured (in cm) at the age of 12 years, and stem sweep (SWE; stem deviation from verticality at 1.5 m from the ground, expressed in cm) was measured at the age of eight years.

### **Breeding value prediction and genetic index**

Breeding values for growth (height and girth) and stem sweep were estimated with the TREEPLAN genetic evaluation system (McRae et al. 2004), which includes a database of all available data from the genetic trials of the French maritime pine breeding program. The phenotypic data were first spatially adjusted within each trial. A joint multivariate analysis of all trials based on the best linear unbiased prediction (BLUP) method was then carried out, taking into account both the pedigree relationships between trees and the correlations between traits. Estimated breeding values (EBVs) were estimated for GBH, HT and SWE at both measurement ages (*i.e.* 8 and 12 years). EBVs were also estimated at harvest age (*i.e.* 35 years), on the basis of age-age correlations for SWE and volume (VOL).

EBVs were obtained with two different pedigree models (before and after pedigree recovery): the partial pedigree (PP) model, in which only the theoretical seed donors were known, and the full-pedigree (FP) model, which included the complete pedigree of the genotyped G2 trees. EBVs are expressed in units of additive standard deviation (with the G0 population as the reference population).

Selection decisions were based on a multiple-trait selection index, combining EBVs as follows:  $\text{Index}_{35} = \text{EBV\_VOL}_{35} - \text{EBV\_SWE}_{35}$ , where  $\text{EBV\_VOL}_{35}$  is the EBV for volume at 35 years of age, and  $\text{EBV\_SWE}_{35}$  is the EBV for stem sweep at 35 years of age.  $\text{Index\_PP}_{35}$  was calculated from the EBVs obtained with the partial pedigree model, and  $\text{Index\_FP}_{35}$  was calculated from the EBVs obtained with the full-pedigree model.

### **Sampling in the polycross trial, with two different preselection strategies**

The G2 trees of the progeny polycross trial were ranked according to Index\_PP\_35. As this index includes no evaluation of major defects, G2 trees were also scored visually (binary score: 0 for trees with major defects, such as bad branching, forks, disease or pest damage; 1 for trees without major defects). Trees with a score of 0 were excluded from the preselection process described below.

Two different options were used in the polycross trial to preselect candidates with high growth and low sweep for pedigree recovery. The two options differed in terms of the contribution of the maternal family:

- In preselection 1 (PS1), no restriction was placed on relatedness. PS1 involved the preselection of trees with no major defects ranked among the 200 best individuals (based on Index\_PP\_35). In total, 153 G2 trees were sampled.
- Preselection 2 (PS2) included a restriction on relatedness. PS2 involved preselection of the two top-ranking trees with no major defects from each of the 75 best families in the progeny trial. The families and the trees within each family were ranked according to Index\_PP\_35. Thus, 150 G2 trees (2 individuals x 75 families) were sampled.

Young needles were collected from the 153 (PS1) and 150 (PS2) preselected trees and their potential parents (seed donors and pollen donors of both polymixes) and stored at -80°C for DNA extraction.

### **DNA extraction and fingerprinting**

Frozen needle tissues were ground to a fine powder and used for DNA extraction with an Invisorb® DNA Plant HTS 96 Kit (Strattec Molecular, Berlin, Germany), according to the manufacturer's instructions. The DNA was quantified with a NanoDrop microvolume spectrophotometer (Thermo Fisher Scientific Inc., Waltham, CA, USA). The sampled individuals were genotyped with SNP markers, in the Sequenom MassARRAY iPLEX Gold assay (Sequenom, San Diego, CA, USA), performed at the genotyping and sequencing facility of Bordeaux, France (<http://www.pgtb.u-bordeaux2.fr/>). The 80 SNPs used here were originally developed for paternity recovery in a maritime pine breeding population (Vidal et al. 2015). These SNPs were selected from a 12k Infinium SNP-array (Illumina, San Diego, USA) developed by Chancerel et al. (2013), and each had a minor allele frequency greater than 0.45 and low levels of linkage disequilibrium ( $r_v^2 < 0.3$ ).

### **Assignment of parentage for the preselected trees**

Likelihood inference was carried out with Cervus 3.0 (Kalinowski et al. 2007; Marshall et al. 1998), both to check the identity of the maternal parent and to recover the identity of the paternal parent for each of the preselected G2 trees. Cervus was run assuming a 0.1% genotyping error rate. The female parent was confirmed if the LOD score was positive and only one mismatch allele was allowed for each progeny and its supposed female parent. For paternity recovery, 90% of the pollen donors were considered to have been sampled (Vidal et al. 2015). The delta score (*i.e.* the difference in LOD scores of the two most likely candidate parents) was used as a criterion for paternity assignment at the 99% confidence level. The critical values of delta scores were based on simulations of 100,000 progeny. One mismatch allele was allowed between a given progeny and its male parent.



### **Final selection for clonal seed orchard establishment**

OPSEL 1.0 software (Mullin 2014), was used for the optimal selection of a production population (virtual CSO), maximizing genetic gains while imposing various constraints on coancestry within the selected population. Constraints on coancestry were based on the minimum status number  $N_s$ . The status number of a population describes the effective number of individuals, *i.e.* the corresponding number of unrelated and non-inbred individuals (Lindgren *et al.* 1997). Three levels of coancestry constraints were tested: either no restriction on  $N_s$ , or  $N_s=10$  or 20. The "optimum selection of seed orchard method" was used, allowing unequal numbers of ramets per genotype in the CSO.

The final selection strategies studied were:

- Forward (FOR) selection, based on preselection PS1 or PS2. The candidate genotypes were G2 trees for which a complete pedigree had been recovered. Genetic evaluation was carried out with Index\_FP\_35 (*i.e.* with EBVs estimated from the full-pedigree model)
- Backward (BACK) selection: all the 166 seed donors (G1) evaluated in the polycross trial were candidates. Genetic evaluation was carried out with Index\_PP\_35 (*i.e.* with EBVs estimated from the partial pedigree model)
- Mixed (MIX) Forward-Backward selection: G2 trees for which a complete pedigree had been recovered and all 166 seed donors were candidates. The genetic evaluation was carried out with Index\_FP\_35 for G1 and G2 individuals.

The target census size of the selected population (total number of ramets constituting the CSO) was set at 600. For logistical reasons, the number of ramets per genotype was set at 50 for G2 trees and 200 for G1 trees (several ramets of G1 trees were available from clonal archives, but this was not the case for G2 trees, with only one tree per genotype, limiting the number of available scions for grafting).

### **Estimation of genetic gain for seed orchards**

The expected genetic gain ( $\Delta G$ ) was calculated as  $\Delta G = CV_a \sum_{i=1}^n EBVi pi$ , where  $CV_a$  is the additive coefficient of variation of the base population (G0 trees),  $EBVi$  and  $pi$  are the estimated breeding value and the proportion of ramets in the CSO of genotype  $i$ , respectively, and  $n$  is the number of different genotypes in the CSO.

$CV_a$  values for height, girth and stem sweep were extracted from the article by Bouffier *et al.* (2008) and were calculated as  $CV_a = \sigma_a/\mu$ , where  $\sigma_a$  is the square root of the additive genetic variance and  $\mu$  is the mean value for the trait. Expected genetic gains are expressed as a percentage relative to G0 trees (plus trees) performances.

## Results

In this study, a breeding strategy was defined as a combination of two selection steps (preselection and final selection), at a given diversity level. Two preselection options (PS1, with no restriction of within-family selection; and PS2, with restriction), three final selection strategies (forward (FOR), backward (BACK) and mixed (MIX) selection), and three diversity levels (no constraint on Ns, Ns=10 and Ns=20) were investigated. The resulting breeding strategies were named according to the combination of these three features. For example, in strategy "FOR\_PS1\_Ns10", forward selection was performed on the candidate trees from preselection option PS1, with a minimum status number of 10 in the CSO.

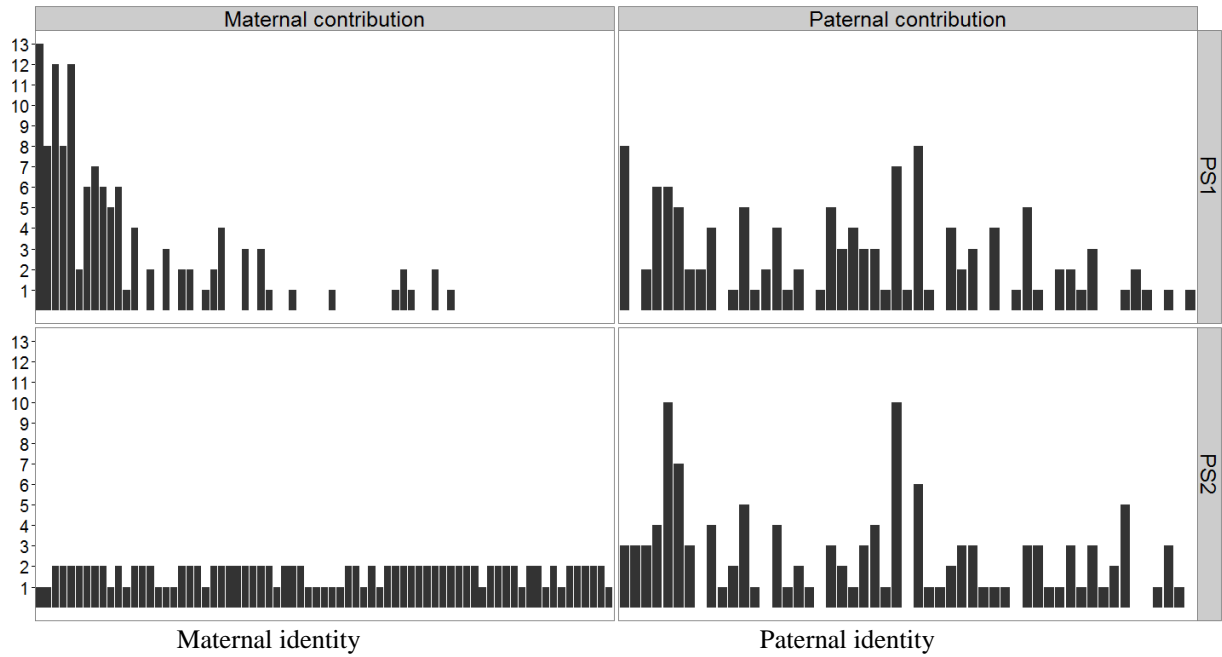
### **Sampling and genotyping**

Two different types of preselection were applied to candidate genotypes in the polycross trial studied: i) PS1 provided 153 G2 trees from 35 half-sib families (with a family size of 1 to 21 individuals/family); ii) PS2 provided 150 G2 trees from 75 half-sib families (2 individuals/family). Overall, 57 individuals were preselected in both approaches. Genotyping was successfully achieved for 146 (PS1) and 147 (PS2) G2 individuals (minimum=45 SNPs, maximum=63 SNPs, mean= 60.5 SNPs), which were analyzed for paternity recovery (Table 1).

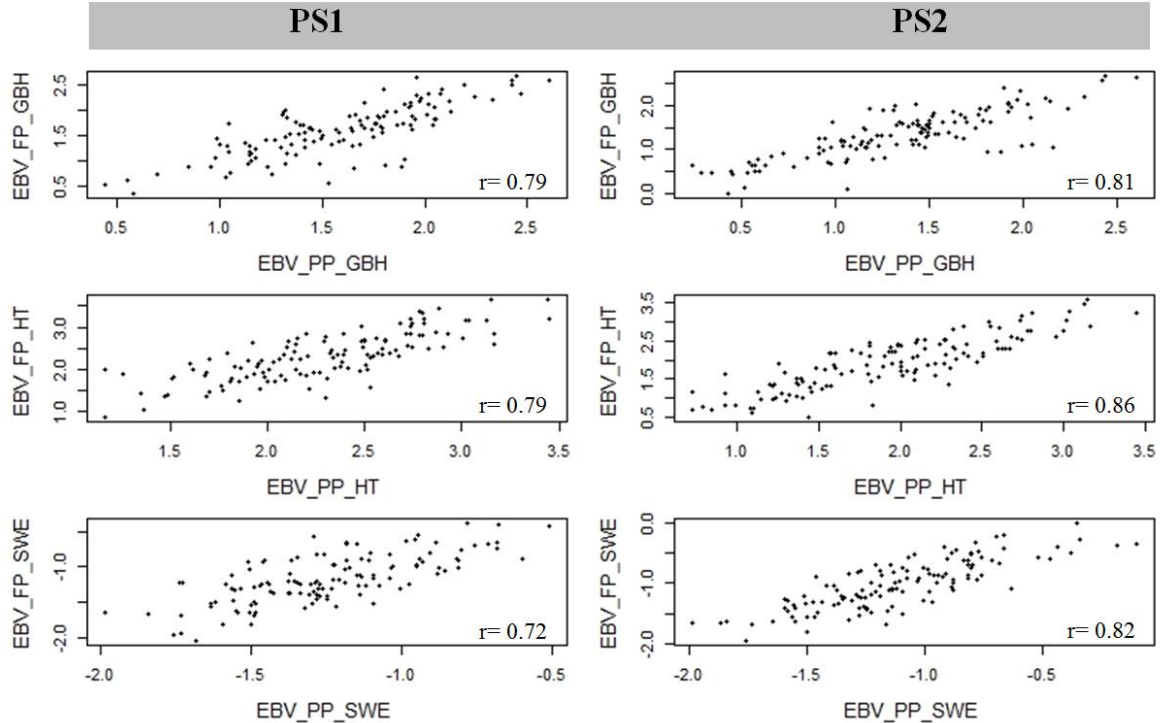
*Table 1 Number of G2 trees sampled with two preselection options (PS1 and PS2) and paternity recovery statistics*

	Preselection PS1	Preselection PS2
Number of preselected G2 trees	153	150
Number of successfully genotyped G2 trees	146	147
Number of G2 trees with seed donor clone confirmed *	145	142
Number of G2 trees with complete pedigree recovered	122	125

\* and thus analyzed for paternity recovery



*Fig. 2 Maternal and paternal contributions (number of progeny per parent) for candidate trees for which a complete pedigree was recovered, for preselection options PS1 and PS2. The identities of the maternal and paternal parents are listed in decreasing order of Index\_PP\_35. A maternal family size of 1 in PS2 means that one of the two preselected G2 individuals has an unknown father (i.e. not included in the polymixes).*



*Fig. 3 Correlation between EBV\_PP (estimated with the partial pedigree model) and EBV\_FP (estimated with full-pedigree model) of the G2 candidate trees preselected with PS1 or PS2 for girth (GBH) and height (HT) at 12 years and for stem sweep (SWE) at eight years, where  $r$  is the Pearson product moment correlation coefficient.*

### **Pedigree recovery on preselected trees**

The identity of the maternal parent was not confirmed for one of the 146 individuals in PS1 and five of the 147 individuals in PS2 analyzed for paternity recovery with Cervus software. These individuals were thus excluded from the paternity analysis. The identity of the maternal parent was confirmed for 122 individuals in PS1 and 125 individuals in PS2, and paternity was recovered with 99% confidence for these individuals (see Table 1). In total, 23 individuals in PS1 and 17 individuals in PS2 respectively, were clearly fathered by outside pollen (*i.e.* not from the two polymixes).

Only individuals for which a complete pedigree was recovered (*i.e.* the mother confirmed and the father identified) were considered as candidate trees for final selection. All subsequent analyses therefore focus on G2 trees with a complete pedigree.

The 122 G2 candidate trees from PS1 came from 30 different seed donor clones (maternal contribution of 1 to 13) and from 42 pollen donor clones (paternal contribution of 1 to 8) (Fig. 2). Mean coancestry within these candidate trees was 0.029 (equivalent to  $N_s=17$ ). The best seed donor clones contributed more than the others (Fig. 2), because the best preselected trees were from the best maternal families.

The 125 G2 candidate trees from PS2 came from 73 different seed donor clones (maternal contribution of 1 to 2) and from 45 pollen donor clones (paternal contribution of 1 to 10) (Fig. 2). Mean coancestry within these candidate trees was 0.017 (equivalent to  $N_s=29$ ).

Overall, 30 seed donors and 34 pollen donors were common to both PS1 and PS2.

### **Correlation between the EBV-PP and EBV-FP of candidate trees**

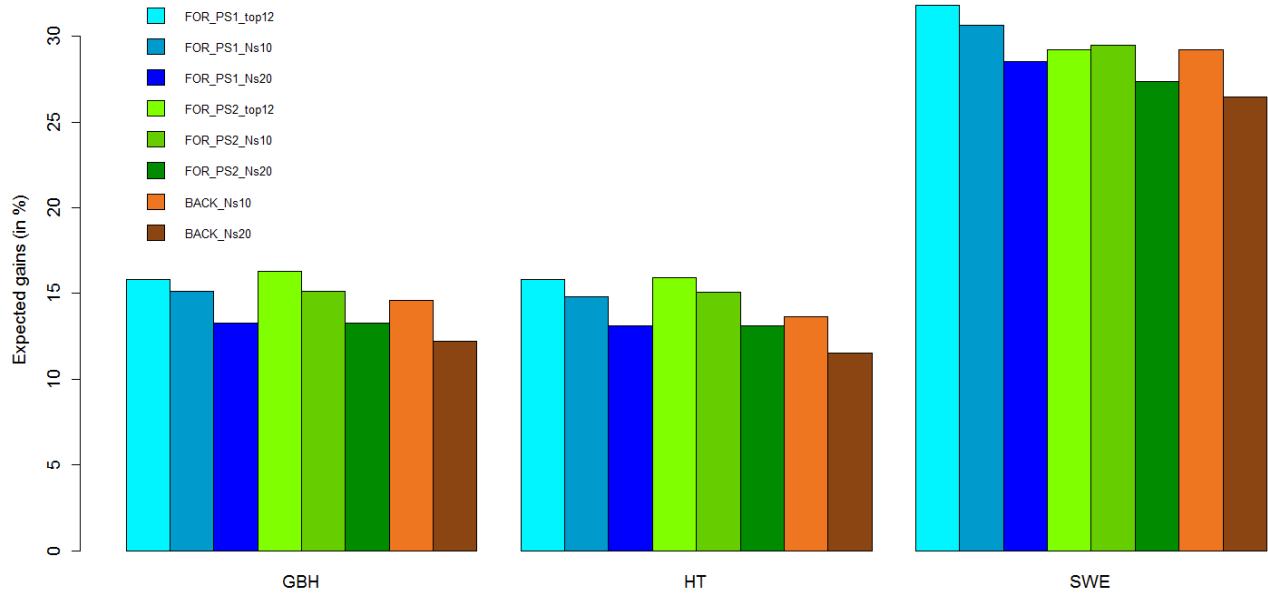
There was a strong correlation between the breeding values estimated with the partial (PP) and full (FP) pedigree models for the three traits in candidate trees from PS1 and PS2. Pearson's correlation coefficients ranged from 0.72 to 0.79 for candidate trees from PS1 and from 0.81 to 0.86 for candidate trees from PS2, depending on the trait considered (Fig. 3). These correlation coefficients were slightly higher in the sample from PS2, probably because the range of EBVs was larger in PS2 than in PS1, due to the limitation of relatedness in PS1.

The accuracy of EBVs estimated with the FP model (0.70 for girth, 0.74 for height, and 0.75 for stem sweep) was much higher than that of EBVs estimated with the PP model (0.56 for girth, 0.63 for height, and 0.65 for stem sweep). As expected, paternity recovery resulted in a better EBV estimation for G2 trees.

### **Final selection to establish clonal seed orchards**

The last step in the selection process was the selection of the best genotypes for which a complete pedigree had been recovered, from the candidate PS1 and PS2 trees, to obtain a virtual CSO. OPSEL software was used to obtain an optimal selection, maximizing genetic gain while maintaining genetic diversity by imposing a constraints on mean relatedness (a minimum status number). The optimization of CSO composition by OPSEL resulted in different numbers of ramets for different genotypes. Genotypes with higher breeding values tended to be represented by larger numbers of ramets, but this trend was counterbalanced by relatedness between these genotypes. Forward selection strategies were evaluated with OPSEL and compared with backward and mixed selection strategies, as explained in the materials and methods section.

5. Evaluation d'une stratégie de sélection forward de type polymix breeding



*Fig. 4 Expected genetic gain (in %, relative to the base population G0) for girth (GBH), height (HT), and stem sweep (SWE; here, positive gains for SWE indicate greater stem straightness) with different breeding strategies: forward selection with PS1 (FOR\_PS1, in blue) or PS2 (FOR\_PS2, in green) and either no restriction on Ns (top12), or Ns=10 or Ns=20; backward selection with Ns constraint (BACK\_Ns10 and BACK\_Ns20, in orange). Selection was optimized with OPSEL software.*

## 5. Evaluation d'une stratégie de sélection forward de type polymix breeding

Forward selection with two preselection options was first studied, with the imposition of different levels of coancestry in the CSO: either no restriction on  $N_s$ , or  $N_s=10$  or  $N_s=20$ . In FOR selection without restriction on  $N_s$ , the 12 best genotypes (Top\_12) were selected (ranking on index\_FP) with either the PS1 or the PS2 option. Each genotype contributed equally to the CSO, with 50 ramets per genotype. PS1 and PS2 gave similar expected gains for height and girth (additional gains of about 16% for HT and GBH, Fig. 4) but PS1 resulted in lower diversity ( $N_s=5$  in strategy FOR\_PS1\_Top12 whereas  $N_s=7$  in strategy FOR\_PS2\_Top12). In FOR selection with constraints on coancestry ( $N_s=10$  or 20), different genotypes contributed different numbers of ramets. As expected, increasing the minimum target status number increased the number of genotypes selected: 37 genotypes (with strategy FOR\_PS1\_Ns10) or 20 genotypes (with strategy FOR\_PS2\_Ns10) contributed to the CSO with 2 to 50 ramets per genotype, whereas 77 (with strategy FOR\_PS1\_Ns20) or 57 (with strategy FOR\_PS2\_Ns20) genotypes contributed to the CSO with 1 to 31 ramets per genotype. The expected genetic gain was decreased slightly with increasing strength of constraint on diversity ( $N_s$ ), regardless of the preselection option (Fig. 4).

Forward selection was then compared with backward selection. All 166 seed donors (G1) evaluated in the polycross trial were candidates for selection on the basis of their Index\_PP (no pedigree recovery for classical backward selection). Forward selection provided a slightly greater genetic gain than backward selection at equivalent  $N_s$  values (Fig. 4). For example, FOR\_PS1\_Ns10 gave an additional gain of 1.4% for SWE, 0.5% for GBH and 1.2% for HT (equivalent to an additional gain of 2.2% for volume) over BACK\_Ns10. However, the expected gain was more reliable in backward selection than in forward selection. The mean EBV accuracy for G1 trees (in backward selection) was about 0.95, whereas that for G2 trees (in forward selection) was about 0.73.

Finally, mixed selection strategies were evaluated. In this case, 37 genotypes (31 G2 and 6 G1) were involved in the CSO for the MIX\_PS1\_Ns10 strategy, and 26 genotypes (19 G2 and 7 G1) were involved in the CSO for the MIX\_PS2\_Ns10 strategy. The two preselection options provided equivalent genetic gains at equivalent  $N_s$  values. Moreover, mixed selection provided gains similar to those achieved with forward selection at equivalent  $N_s$  values. For example, the MIX\_PS1\_Ns10 strategy yielded an additional gain of 1.4% for SWE, 0.3% for GBH and 0.2% for HT (equivalent to an additional gain of 0.8% for volume) over the FOR\_PS1\_Ns10 strategy.

## Supplemental table: summary of breeding strategies

Estimation of expected genetic gains for girth (GBH), height (HT) and stem sweep (SWE) in virtual clonal seed orchards (CSO) established from different breeding strategies: forward selections (from preselection strategies S1 or S2) with either no restriction on Ns (top12), or Ns=10 or Ns=20; backward selection with constraint on Ns (10 or 20); mixed selections with constraint on Ns (10 or 20). Optimization of selection was done with OPSEL software. The target census size of the selected population (total number of ramets constituting the CSO) was set at 600. The number of ramets available per genotype was set at 50 for G2 trees and 200 for G1 trees

Final selection strategy	Forward Top12 G2		Forward with restriction on Ns				Backward with restriction on Ns		Mixed (FOR_BACK) with restriction on Ns							
	NO restriction		NS=10		NS=20		NS=10		NS=20		NS=10		NS=20			
	PS1	PS2	PS1	PS2	PS1	PS2	PS1	PS2	PS1	PS2	PS1	PS2	PS1	PS2		
Restriction on status number (Ns)	12 best G2		122 G2		122 G2		166 G1		166 G1		122 G2 + 166 G1		122 G2 + 166 G1		125 G2 + 166 G1	
Preselection option	12 best G2		122 G2		122 G2		166 G1		166 G1		122 G2 + 166 G1		122 G2 + 166 G1		125 G2 + 166 G1	
Number of genotypes available	12		37		77		23		43		37 (31 G2 and 6 G1)		26 (19 G2 and 7 G1)		77 (58 G2 and 19 G1)	
Number of genotype in the CSO	50 (50-50)		16.2 (2-49)		7.8 (1-22)		26.1 (1-113)		13.9 (1-55)		16.3 (1-107)		23.1 (1-114)		7.8 (1-39)	
Number of ramets/genotype (mean(min-max))	3.3		3.1		2.8		3.0		2.6		3.2		3.2		2.9	
Average index of the CSO	3.2		3.1		2.8		3.0		2.6		3.2		3.2		2.9	
Genetic gain for GBH at age 12 (in %)	15.8		15.1		13.3		14.6		12.2		15.4		15.7		13.5	
Genetic gain for HT at age 12 (in %)	15.8		14.8		13.1		13.6		11.6		15.0		15.2		13.2	
Genetic gain for SWE at age 8 (in %)	31.8		30.6		28.5		29.2		26.4		32.0		31.6		29.2	

## Discussion

The main objective of this study was to assess the feasibility of forward selection associated with parental analysis of the progeny in an existing maritime pine polycross progeny trial. A few theoretical studies have been carried out but, to our knowledge, this is the first example of a practical study of forward selection in a polycross trial. Different options for the forward and classical backward selection of a production population (establishment of a virtual CSO) were studied and compared on the basis of genetic gains for growth traits and stem straightness.

The various stages of forward selection, and some considerations about the PMX/WPA strategy, are discussed below.

### ***Preselection options in a polycross trial***

The genotyping of all individuals in a progeny trial would be too costly, so the preselection of trees is a necessary step before paternity recovery. This step must provide candidate trees for the final selection with two goals: maximizing genetic gain whilst limiting the relatedness between candidates to ensure that the CSO population contains sufficient diversity.

The identities of the pollen donors were unknown. Consequently, one limitation of this approach was that the set of preselected candidates may not have included some of the best individuals from the polycross trial due to inaccurate EBV estimations (obtained with the PP model). Nevertheless, we showed that the between EBVs estimated with the partial and full-pedigree models were highly correlated. The ranking of G2 trees on the basis of their index\_FP would therefore have been relatively similar to that obtained with index\_PP if pedigrees had been determined for all the trees.

The preselection of candidate trees may affect final genetic gain and diversity in the CSO. We therefore considered two contrasting preselection options, one with (PS2) and the other without (PS1) restrictions on relatedness between the preselected candidates. In this study, the choice of PS1 or PS2 had little effect on the final selection, as these options yielded similar genetic gains at equivalent  $N_s$  values, mostly because the number of preselected individuals was high, and the bias in the EBVs estimated with the PP model was small. The same number of individuals was sampled in PS1 and PS2, but the mean coancestry (calculated with complete pedigree information) was, as expected, higher for PS1 than for PS2. Thus, for equivalent genetic gain and diversity in the CSO, PS1 resulted in the selection of a larger number of different genotypes, with fewer ramets per clone required than PS2. PS1 was therefore more logistically efficient, as fewer scions per tree were required. Thus, PS1 seems to be the most appropriate preselection approach for our breeding program, and it does not seem to be necessary to apply constraints on relatedness between preselected individuals, provided that enough trees are preselected.

### ***Genetic gain and diversity in commercial seed orchards***

A large proportion of the planting material for cultivated forests today originates from seed orchards. For maritime pine, more than 90% of the seedlings used for the reforestation of the Landes in Gascony are improved seedlings originating from seed orchards (GIS PMF 2014). Seed orchards consist of selected superior individuals, and the main objective of their establishment is to generate genetically improved forest tree seeds by maximizing genetic gain (Funda and El-Kassaby (2012)). The challenge for tree breeders is thus to create seed orchards in which breeding progress is maximal (maximum performance), but with a sufficient degree of genetic diversity to ensure a reasonable degree of genetic heterogeneity in the final forest (Hosius et al. 2000; Lindgren et al. 2009; Stoehr et



## 5. Evaluation d'une stratégie de sélection forward de type polymix breeding

al. 2004). Genetic diversity plays an important role in the sustainability of forest ecosystems and is essential for a population to adapt to new environmental factors, such as climate change and diseases (Hansen 2008; Johnson and Lipow 2002; Muller-Starck 1995). How much genetic diversity should be present in a CSO depends on the length of the rotation and the environmental variation to which the planting material originating from the CSO will need to adapt during its lifetime (Johnson and Lipow 2002). Johnson and Lipow showed that a seed orchard with "25 unrelated selections contains about 92 percent of the genetic variation of the natural population" and that a minimum of "20 unrelated selections should provide the same level of risks as seed collected from the natural population". Moreover, restrictions on relatedness between the individuals selected for the CSO can limit inbreeding depression, with potential effects on the performance of the planting material (Durel et al. 1996; Olsson et al. 2001; Stoehr et al. 2008). However, the management of diversity and relatedness (expressed as group coancestry here) between selections becomes relatively complicated at the 3<sup>rd</sup> generation of breeding. OPSEL software proposes an optimal selection, "not to completely avoid kinship, but rather to find the set of selections that maximizes gain under a relatedness constraint" (Mullin 2014). In this study, we used the status number  $N_s$  to quantify coancestry in the CSO.  $N_s$  is a useful parameter for evaluating trade-offs between gain and diversity (Lindgren et al. 1997; Lindgren and Kang 1997; Lindgren and Mullin 1998). The minimum  $N_s$  was set at 10 or 20 for a population census size of 600 (total number of ramets in the CSO).

In the French maritime pine breeding program, the establishment of a CSO based on forward selection could involve the best G2 trees from several polycross progeny trials. Our standard CSO area is at least 10 ha, so about 2,400 grafted trees would be required. Four polycross trials are currently available for forward selection, so the selection of the best genotypes providing 600 ramets within a trial was considered here. Due to the relatedness between the parents used in the different polycross trials, we set the diversity in the studied trial at  $N_s=10$ , to ensure that the minimum diversity required was attained (as described above and in accordance with (Johnson and Lipow 2002) in the final complete CSO.

The number of ramets per genotype is limited in forward selection approaches, because each selected clone is represented by a single tree (giving few scions). By contrast, in backward selection, the parent trees selected are often grafted with several replicates in clonal archives. The development of efficient vegetative propagation methods (such as micropropagation, somatic embryogenesis or micrografting through tissue culture) would increase the number of ramets available for the best genotypes and provide powerful tools for scaling up the production of genetically improved planting material (Bonga 2015; Lelu-Walter et al. 2013). However, such methods are not yet available for use in this species (and were therefore not considered in our options).

Whatever the preselection option used, forward selection resulted in a greater genetic gain than backward selection. It should be borne in mind that the polycross trial studied was not designed with forward selection in mind. In particular, the pollen mixes were mostly of random composition rather than based on high EBVs. The expected genetic gain obtained with forward selection in this trial would therefore be far from optimal.

Finally, genetic gain and diversity in the production population were estimated under an assumption of random mating, equal reproductive success and no pollen contamination within the CSO. However, many factors can affect the genetic quality of orchard seed-lots. Both genetic gain and diversity depend on the variation of reproductive success in the CSO, synchrony in reproductive phenology, pollen quality and contamination, self-fertilization rates, seed germination, and other factors. Many studies have shown that there can be a considerable gulf between expectations and

## 5. Evaluation d'une stratégie de sélection forward de type polymix breeding

reality (Askew 1988; Burczyk et al. 1997; Edwards and ElKassaby 1996; Funda et al. 2009; Gomory et al. 2003; Hansen 2008; Kang and Lindgren 1998; Machanska et al. 2013; Matziris 1994; Na et al. 2015), making it difficult to predict genetic quality. Moreover, the selfing rate in the CSO and inbreeding depression were not taken into account in the estimation of genetic gain and diversity. However, absolute values were not of prime importance here as the aim was to compare different breeding strategies.

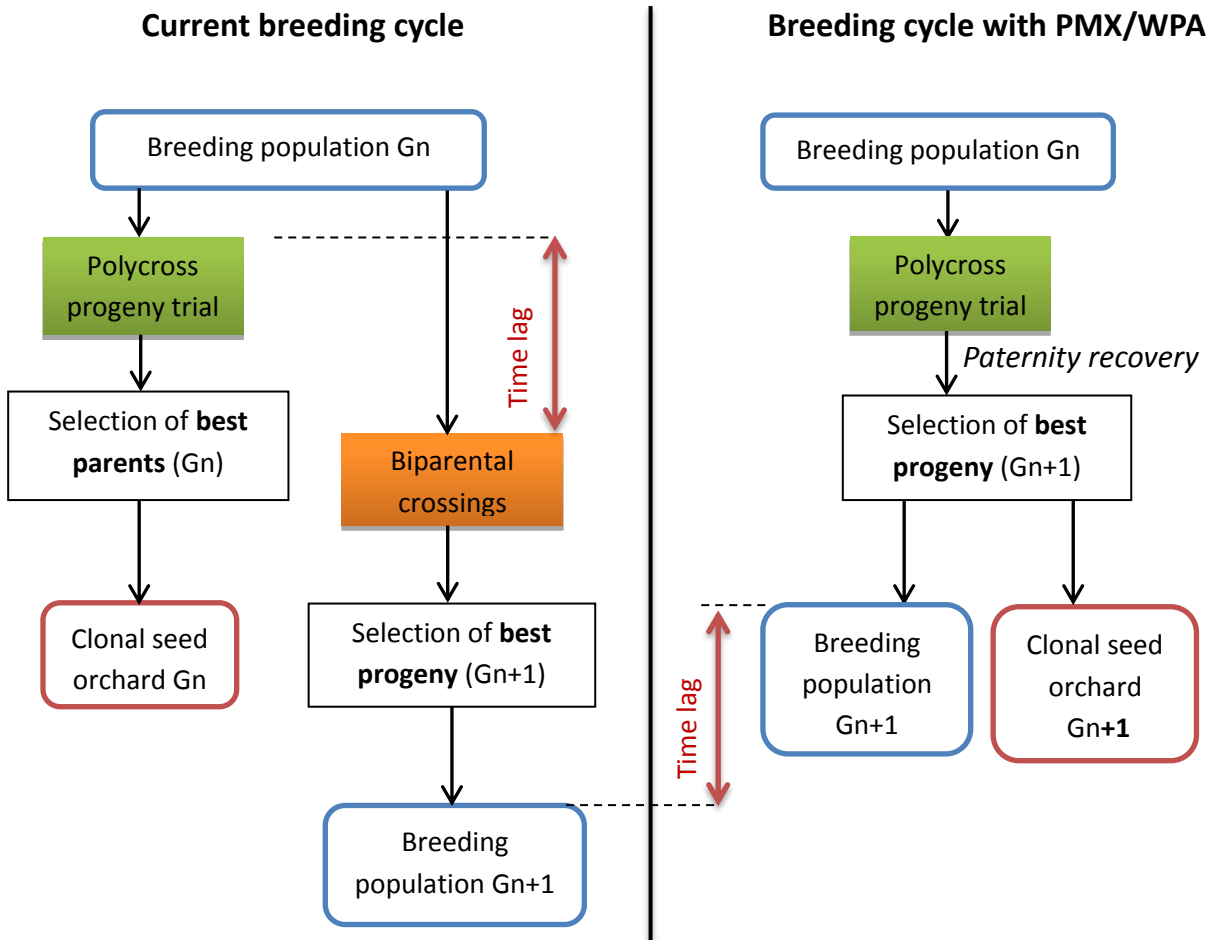
### **Towards the implementation of a PMX/WPA strategy?**

This study shows that forward selection associated with molecular pedigree analysis of progeny is feasible in real-life conditions. In the polycross trial analyzed here, the expected genetic gain in the production population did not much exceed that obtained with classical backward selection, largely because this polycross trial was not designed for this kind of selection.

The implementation of a PMX/WPA strategy, as proposed by Lambeth et al. (2001) is a broader issue, because the best progeny are selected for the next generation of breeding, rather than just for a CSO. This strategy has clear advantages, including the need for only one round of crossing (polycross), for simultaneous testing and recruitment for forward selection. The many costly full-sib crosses required for classical approaches are replaced by a small number of polycrosses (with one or several different polymixes), followed by genotyping and paternity recovery to identify the best progeny for use in the next generation of breeding. This approach is thus easier to implement than classical approaches, and, as breeding and testing are performed at the same time, the interval between the generations of consecutive seed orchards is shortened (Fig. 5). In our current breeding cycle, because of technical and economic constraints (number of crosses possible in each year) and because the results of polycross trials are used to choose parents for full-sib crosses, there is a time lag between polymix crosses (for parental EBV estimation) and full-sib crosses (used to select the next generation for breeding). The use of a PMX/WPA strategy would eliminate this time lag (Fig. 5), accelerating the breeding cycle. Improved planting material can be renewed more quickly and adapts better to changing economic and environmental contexts.

Polycrosses also maximize the number of full-sib families obtained with a smaller number of crosses than for full-sib designs. However, the combinations of parents are not precisely chosen. The parents transferring their genes to the next breeding generation are therefore determined, to some extent, at the time of selection, but not at the time of the cross. However, by contrast to the "breeding without breeding" approach (pedigree identification from open pollination rather than controlled crosses), the potential fathers are at least partially selected through the choice of polymix composition.

The successful use of polymix breeding and testing systems requires accurate pedigree reconstruction, small differences in male reproductive success (to prevent the difficulties involved in managing coancestry within the preselected subset in situations in which the best fathers contribute more to the progeny than others), and a low rate of pollen contamination (because only trees with a full pedigree can be selected). Vidal *et al.* (2015) recently showed that these requirements are fulfilled in the maritime pine polycross progeny trial studied here. Additional studies are required, to optimize the polymix trial design (e.g. composition and number of polymixes) in particular. A relatively large number of pollen donors with low levels of relatedness and high EBVs is required within the polymix, to ensure that there will be sufficient genetic diversity in the next generation of the breeding population.



*Fig. 5 Comparison of the current breeding cycle based on backward selection and a breeding cycle based on forward selection as in a “polymix breeding with parental analysis” (PMX/WPA) strategy. A breeding cycle based on a PMX/WPA strategy runs faster because it eliminates the time lag between polymix crossing and biparental crossing.  $G_n$  and  $G_{n+1}$  are generations  $n$  and  $n+1$  of the breeding population, respectively.*

## Conclusion

This study shows that forward selection with pedigree reconstruction is feasible for maritime pine. Complete pedigrees were recovered for most of the preselected (and thus genotyped) progeny, a prerequisite for selection for the production population. In the polycross progeny trial analyzed, forward selection gave a slightly greater genetic gain (despite the absence of optimization of polymix composition) than classical backward selection. No major differences in expected genetic gain in the production population were observed between two contrasting preselection options (with and without constraints on relatedness).

The implementation of a PMX/WPA strategy, speeding up the production of the next breeding population and decreasing the workload, would be possible. However, simulation studies are required to optimize the general design of such breeding strategies, and a cost/benefit analysis should be performed to assess their economic efficiency, given the specific cost and time components of our maritime pine breeding programme.

Forward selection also provides a favorable context for genomic selection. Indeed, the additive genetic relationship matrix (derived from the pedigree) could be replaced with a genomic relationship matrix, to improve the estimation of EBVs and to ensure the maintenance of higher levels of genetic diversity within the breeding program.

## Contribution of the co-authors

MV sampled plant material and extracted DNA. MV and LB analyzed the data. MV wrote the manuscript, helped by AR, CP and LB. AR, LB, CP and LH read and revised the manuscript. LB designed and coordinated the study. All authors read and approved the final manuscript.

## Funding

This study was funded by INRA (EFPA division " *projet innovant* "), the European-Union (ProCoGen project: n°289841) and *Conseil Régional d'Aquitaine* (IMAF project co-funded by FCBA: n° 120009468-052). Marjorie Vidal received a CIFRE PhD fellowship (Public/Private Research Partnerships between FCBA and the French Ministry of Higher Education and Research).

## Acknowledgements

This study would not have been possible without the support of the maritime Pine Breeding Cooperative (GIS " *Pin Maritime du Futur* "). We gratefully acknowledge all its members. The authors also thank the *INRA Experimental Unit (UE0570)* for field measurements, Jean-Mathieu De Boisseson (FCBA) for needles sampling, Tim Mullin for providing access to OPSEL Software, Jérôme Bartholomé (INRA) for useful advices in R.

The genotyping was performed at the Genomic Facility of Bordeaux (grants from the *Conseil Régional d'Aquitaine*: n°20030304002FA and 20040305003FA, the European Union: FEDER n°2003227 and ANR: n°ANR-10-EQPX-16 Xyloforest) with help from Christophe Boury and Adline Delcamp (INRA).

## References

- Askew GR (1988) Estimation of gamete pool compositions in clonal seed orchards. *Silvae Genet* 37:227-232
- Bonga JM (2015) A comparative evaluation of the application of somatic embryogenesis, rooting of cuttings, and organogenesis of conifers. *Can J For Res* 45:379–383. doi:dx.doi.org/10.1139/cjfr-2014-0360
- Bouffier L, Raffin A, Kremer A (2008) Evolution of genetic variation for selected traits in successive breeding populations of maritime pine. *Heredity* 101:156-165. doi:10.1038/hdy.2008.41
- Burczyk J, Nikkanen T, Lewandowski A (1997) Evidence of an unbalanced mating pattern in a seed orchard composed of two larch species. *Silvae Genet* 46:176-181
- Burdon RD, Kumar S (2004) Forwards versus backwards selection: trade-offs between expected genetic gain and risk avoidance. *N Z J Forest Sci* 34:3-21
- Chancerel E, Lamy JB, Lesur I, Noirot C, Klopp C, Ehrenmann F, Boury C, Le Provost G, Label P, Lalanne C, Leger V, Salin F, Gion JM, Plomion C (2013) High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination. *BMC Biol* 11:19. doi:10.1186/1741-7007-11-50
- Cornelius J (1994) Heritabilities and additive genetic coefficients of variation in forest trees. *Can J For Res* 24:372-379. doi:10.1139/x94-050
- Durel CE, Bertin P, Kremer A (1996) Relationship between inbreeding depression and inbreeding coefficient in maritime pine (*Pinus pinaster*). *Theor Appl Genet* 92:347-356
- Edwards DGW, El-Kassaby YA (1996) The biology and management of coniferous forest seeds: genetic perspectives. *For Chron* 72:481-484
- El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klapste J, Lstiburek M (2011) Breeding without Breeding: Is a Complete Pedigree Necessary for Efficient Breeding? *PLoS One* 6:11. doi:10.1371/journal.pone.0025737
- El-Kassaby YA, Lstiburek M (2009) Breeding without breeding. *Genet Res* 91:111-120. doi:10.1017/s001667230900007x
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. vol Ed. 4. Longman Group Limited, Harlow, UK
- Funda T, El-Kassaby YA (2012) Seed orchard genetics. *CAB Reviews* 7:1-23. doi:10.1079/pavsnr20127013
- Funda T, Lstiburek M, Lachout P, Klapste J, El-Kassaby YA (2009) Optimization of combined genetic gain and diversity for collection and deployment of seed orchard crops. *Tree Genet Genomes* 5:583-593. doi:10.1007/s11295-009-0211-3
- GIS PMF (2014) GIS Groupe Pin maritime du futur. Les cahiers de la reconstitution n°4: matériel végétal de reboisement. <http://www.onf.fr/outils/medias/20130708-143100-661300/++files++/4>.
- Gomory D, Bruchanik R, Longauer R (2003) Fertility variation and flowering asynchrony in *Pinus sylvestris*: consequences for the genetic structure of progeny in seed orchards. *For Ecol Manage* 174:117-126. doi:10.1016/s0378-1127(02)00031-2
- Hansen OK (2008) Mating patterns, genetic composition and diversity levels in two seed orchards with few clones - Impact on planting crop. *For Ecol Manage* 256:1167-1177. doi:10.1016/j.foreco.2008.06.032
- Hosius B, Bergmann F, Konnert M, Henkel W (2000) A concept for seed orchards based on isoenzyme gene markers. *For Ecol Manage* 131:143-152. doi:10.1016/s0378-1127(99)00209-1

## 5. Evaluation d'une stratégie de sélection forward de type polymix breeding

- Illy G (1966) Recherches sur l'amélioration génétique du pin maritime. *Ann Sci forest* 23:765-948. doi:10.1051/forest/19660401
- Johnson R, Lipow S (2002) Compatibility of breeding for increased wood production and long-term sustainability: the genetic variation of seed orchard seed and associated risks. In: Johnson AC, Haynes RW, Monserud RA (eds) *Congruent Management of Multiple Resources: Proceedings from the Wood Compatibility Initiative Workshop*, vol 563. Usda Forest Service General Technical Report Pacific Northwest. Us Dept Agr, Forest Serv Pacific Nw Research Stn, Portland, pp 169-179.
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16:1099-1106. doi:10.1111/j.1365-294X.2007.03089.x
- Kang KS, Lindgren D (1998) Fertility variation and its effect on the relatedness of seeds in *Pinus densiflora*, *Pinus thunbergii* and *Pinus koraiensis* clonal seed orchards. *Silvae Genet* 47:196-201
- Lambeth C, Lee BC, O'Malley D, Wheeler N (2001) Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet* 103:930-943. doi:10.1007/s001220100627
- Lelu-Walter MA, Thompson D, Harvengt L, Sanchez L, Toribio M, Paques LE (2013) Somatic embryogenesis in forestry with a focus on Europe: state-of-the-art, benefits, challenges and future direction. *Tree Genet Genomes* 9:883-899. doi:10.1007/s11295-013-0620-1
- Lindgren D, Danusevicius D, Rosvall O (2009) Unequal deployment of clones to seed orchards by considering genetic gain, relatedness and gene diversity. *Forestry* 82:17-28. doi:10.1093/forestry/cpn033
- Lindgren D, Gea LD, Jefferson PA (1997) Status number for measuring genetic diversity. *Forest Genetics* 4:69-76
- Lindgren D, Kang K (1997) Status number - a useful tool for tree breeding. *Research Report of the Forest Genetics Research Institute (Suwon)*:154-165
- Lindgren D, Mullin TJ (1998) Relatedness and status number in seed orchard crops. *Can J For Res-Rev Can Rech For* 28:276-283. doi:10.1139/cjfr-28-2-276
- Lstiburek M, Hodge GR, Lachout P (2015) Uncovering genetic information from commercial forest plantations-making up for lost time using "Breeding without Breeding". *Tree Genet Genomes* 11:12. doi:10.1007/s11295-015-0881-y
- Lstiburek M, Ivankova K, Kadlec J, Kobliha J, Klapste J, El-Kassaby YA (2011) Breeding without breeding: minimum fingerprinting effort with respect to the effective population size. *Tree Genet Genomes* 7:1069-1078. doi:10.1007/s11295-011-0395-1
- Machanska E, Bajcar V, Longauer R, Gomory D (2013) Effective population size estimation in seed orchards: a case study of *Pinus nigra* ARNOLD and *Fraxinus excelsior* L./ *F. angustifolia* VAHL. *Genetika-Belgrade* 45:575-588. doi:10.2298/genstr1302575m
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7:639-655. doi:10.1046/j.1365-294x.1998.00374.x
- Matziris DI (1994) Genetic variation in the phenology of flowering in black pine. *Silvae Genet* 43:321-328
- McRae TA, Dutkowski GW, Pilbeam DJ, Powell MB, Tier B (2004) Genetic Evaluation Using the TREEPLAN® System. Paper presented at the IUFRO Joint Conference of Division 2 "Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future" Charleston, SC, USA, 1-5 November 2004
- Muller-Starck G (1995) Protection of genetic variability in forest trees. *Forest Genetics* 2:121-124

## 5. Evaluation d'une stratégie de sélection forward de type polymix breeding

- Mullin TJ (2014) OPSEL 1.0: A computer program for optimal selection in forest tree breeding. Technical Report Nr 841-2014, Arbetsrapport Från Skogforsk.
- Na SJ, Lee HS, Han SU, Park JM, Kang KS (2015) Estimation of genetic gain and diversity under various genetic thinning scenarios in a breeding seed orchard of *Quercus acutissima*. *Scand J Forest Res* 30:377-381. doi:10.1080/02827581.2015.1018936
- Olsson T, Lindgren D, Li B (2001) Balancing genetic gain and relatedness in seed orchards. *Silvae Genet* 50:222-227
- Pâques L (2013) Forest tree breeding in Europe: current state-of-the-art and perspectives. ed. / Luc E Pâques. Springer (Managing Forest Ecosystems, Vol. 25). doi:10.1007/978-94-007-6146-9
- Stoehr M, Webber J, Woods J (2004) Protocol for rating seed orchard seedlots in British Columbia: quantifying genetic gain and diversity. *Forestry* 77:297-303. doi:10.1093/forestry/77.4.297
- Stoehr M, Yanchuk A, Xie CY, Sanchez L (2008) Gain and diversity in advanced generation coastal Douglas-fir selections for seed production populations. *Tree Genet Genomes* 4:193-200. doi:10.1007/s11295-007-0100-6
- Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L (2015) Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. *Tree Genet Genomes* 11:1-13. doi: 10.1007/s11295-015-0932-4
- Zobel BJ, Talbert JT (1984) *Applied Forest Tree Improvement* (p. 528). New York: Wiley.

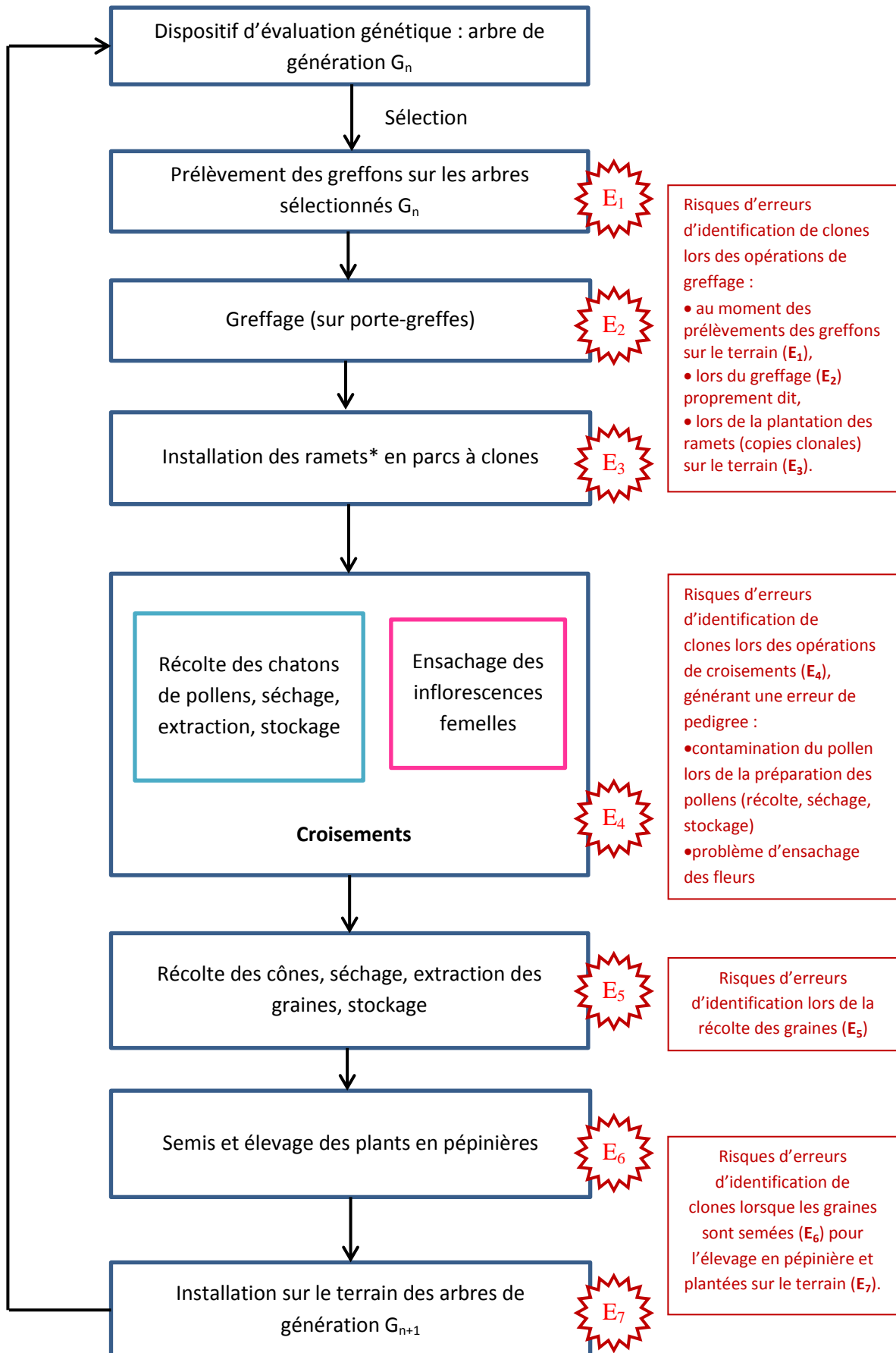
### 5.3 Conclusion de la partie 5

L'**objectif III** consistait à tester la faisabilité d'une stratégie de sélection innovante en réalisant une sélection *forward* de type « *polymix breeding* » au sein d'un test de descendance polycross. On a montré qu'il est possible de sélectionner les meilleurs descendants issus de croisements polycross, dont on aura au préalable reconstitué les pedigrees par génotypage. Ces individus peuvent constituer une population de production via leur installation en verger à graines. La nouvelle variété améliorée issue de ces individus sélectionnés en *forward* présente un gain génétique légèrement supérieur à une variété issue d'individus sélectionnés en *backward*. De plus, la sélection d'individus au sein des tests de descendance polycross peut, à gain équivalent, apporter de la diversité au sein des populations de production. Cette valorisation des tests polycross existants est discutée dans la [partie 6.2](#).

De plus, ce travail ouvre une perspective vers la mise en place d'une stratégie de sélection *forward* au sein du programme d'amélioration du pin maritime, afin d'optimiser les stratégies de sélection au sein du programme d'amélioration génétique et de renouveler les variétés plus fréquemment. Cette perspective est développée dans la [partie 6.3](#).



Figure 26 : Exemples de sources d'erreurs potentielles (notées  $E_n$ ) pouvant compromettre le suivi de l'identité des clones au cours des opérations effectuées dans un cycle de sélection



\*ramets : copies clonales obtenues par voie végétative (généralement par greffage) à partir d'un seul individu originel (appelé ortet).

## 6. Applications et perspectives

Cette dernière partie décrit les principales applications qui découlent de cette thèse menée dans le cadre du programme d'amélioration génétique du pin maritime, à court et long terme.

Dans un premier temps, je présente l'utilisation du jeu de SNPs développé au cours de ce travail. En effet, il est actuellement utilisé pour vérifier les identités génétiques d'individus mais aussi pour estimer le taux de pollution pollinique dans les vergers à graines. D'autres jeux de SNPs pourraient être développés afin de certifier l'origine des matériels de reboisement. Dans un second temps, je discute de la valorisation possible des tests de descendance polycross existants en y réalisant de la sélection *forward* (selon la méthodologie décrite dans la *partie 5*). Dans un troisième temps, je présente l'intérêt de la mise en place d'une stratégie de sélection *forward* au sein du programme d'amélioration. Enfin, je discute des nouvelles perspectives qu'offrent les marqueurs moléculaires et la reconstitution de pedigree, et notamment une possible transition vers la sélection génomique.

### 6.1 Les premières utilisations des marqueurs moléculaires dans le programme d'amélioration du pin maritime

#### 6.1.1 Utilisation du jeu de SNPs développé pendant la thèse

Deux multiplex de 40 SNPs chacun ont été mis au point dans cette étude pour reconstituer le pedigree au sein des populations d'amélioration du pin maritime landais, avec un coût de génotypage relativement faible (~7€/ échantillon). Ils ont permis de vérifier l'identité de la mère et de retrouver le père (parmi les pères potentiels constituant le polymix) d'individus G2 issus de croisements polycross.

Ce panel de SNPs est actuellement valorisé pour deux autres applications dans le cadre du programme d'amélioration du pin maritime.

##### ➤ Vérification des identités

Au cours de ce travail de thèse, différentes erreurs d'identification de clones ont été mises en évidence (erreur de ramets dans les parcs à clones, erreur de pedigree, *etc.*). Plusieurs étapes au cours du programme d'amélioration sont susceptibles de générer des erreurs pouvant compromettre le suivi de l'identité des clones (Figure 26).

Par exemple, lors de l'étape de greffage des individus d'intérêt en parcs à clones, il peut se produire des erreurs d'identification de clones :

- au moment des prélèvements des greffons sur le terrain ( $E_1$ ),
- lors du greffage ( $E_2$ ) proprement dit,
- lors de la plantation des ramets (copies clonales) sur le terrain ( $E_3$ ).

Les opérations de croisements ( $E_4$ ) sont aussi des sources potentielles d'erreurs. En effet, il faut d'une part, récolter les chatons (inflorescences mâles) contenant le pollen, les sécher, en extraire le pollen, et le stocker au froid. Ces étapes peuvent être l'occasion de contamination d'un pollen par un autre, ou de mauvais étiquetage des pollens. D'autre part, les fleurs femelles sont ensachées, pollinisées, et étiquetées. Or, un mauvais ensachage

(mauvaise isolation des fleurs, ensachage trop tardif) peut être une source de contamination par du pollen présent dans l'air.

Ensuite, 20 mois environ après la pollinisation, les cônes sont prélevés, séchés et les graines en sont extraites (E<sub>5</sub>) : il existe alors un risque de récolter des cônes issus de pollinisation libre et non de croisements contrôlés, et aussi un risque de mélange des graines lors de leur extraction des cônes, ou même lors du stockage.

Enfin, les graines sont semées (~ 6 mois plus tard), puis les plants sont élevés en pépinière (E<sub>6</sub>) et plantés sur le terrain (E<sub>7</sub>). Des erreurs d'étiquetage peuvent également se glisser au cours de ces travaux.

Ces difficultés et aléas d'identification des clones au sein des programmes d'amélioration ont été relevés par plusieurs auteurs (Adams *et al.* 1988 ; Doerksen et Herbing 2008 ; Kumar et Richardson 2005 ; Zhao *et al.* 2013). Les erreurs de pedigree et/ou d'identification diminuent la précision de l'estimation des valeurs génétiques et ont un impact sur le gain génétique (Banos *et al.* 2001 ; Israel et Weller 2000 ; Munoz *et al.* 2014 ; Sanders *et al.* 2006). L'utilisation des marqueurs moléculaires permet de vérifier les identités et de corriger les bases de données sur lesquelles reposent les analyses génétiques.

Le jeu de SNPs mis au point dans cette étude est ainsi actuellement utilisé dans le cadre du nouveau programme d'activité 2015-2020 du GIS PMF nommé « *Pinaster* », qui prévoit la vérification des identités de 200 génotypes (greffés et conservés dans les parcs à clones) par an (sur 5 ans) afin de vérifier l'identité des ramets sur le terrain et d'améliorer la traçabilité des populations d'amélioration. Cette vérification est importante car la majorité des croisements sont réalisés dans ces parcs à clones. C'est la descendance de ces croisements qui constitue à la fois les dispositifs d'évaluation génétique, les différentes générations des populations d'amélioration, et même les vergers à graines dits de famille. D'autre part, l'installation des vergers dits de clones est réalisée à partir de greffons prélevés sur les ramets de ces parcs à clones. Il est donc primordial que les identités de ces ramets soient vérifiées.

Cependant, une fois les erreurs détectées, la correction à apporter est loin d'être évidente car l'antériorité et l'origine des erreurs sont généralement difficiles à déterminer. De plus, il y a peu de traçabilité sur la position des ramets utilisés dans les parcs à clones lors des étapes de croisements contrôlés. Par exemple, le pollen est souvent récolté sur plusieurs ramets différents d'un même clone, et ensuite mélangé. Lors de la pollinisation, plusieurs inflorescences femelles de divers ramets sont utilisées pour les croisements contrôlés. Lors de la récolte des graines (issues de ces croisements), on perd généralement l'information indiquant de quel ramet provient quelle graine (position dans le parc à clones). Le problème se complexifie encore lorsque l'on prend en compte l'histoire des individus greffés en parcs à clones. Par exemple, une erreur de pedigree d'un individu G2 à cause d'une erreur d'identité de l'un de ses parents peut avoir une multitude d'origines, et l'enquête peut être très longue à mener. Ainsi, la gestion des erreurs et leur purge de l'ensemble des générations du programme d'amélioration est complexe, mais ce travail permettra la mise à jour des bases de données et augmentera la précision des évaluations génétiques.

#### ➤ **Etude du taux de pollution pollinique dans les vergers à graines**

Le jeu de SNPs développé dans cette thèse est aussi en cours d'utilisation pour estimer le taux de pollution pollinique de différents vergers à graines. L'objectif de cette étude (menée conjointement par l'INRA et FCBA) est d'évaluer l'influence de 4 paramètres sur le taux de

pollution : la localisation du verger (au sein *versus* en dehors du massif landais), la zone du verger (central *versus* périphérique), le génotype de la mère (comparaison de génotypes de phénologie contrastée) et l'année de récolte des graines améliorées. Cette étude vise aussi à préciser le taux d'autofécondation et le succès reproductif des génotypes paternels.

Ces analyses permettront de collecter des informations utiles pour la mise en place et la gestion des futurs vergers.

### 6.1.2 Des SNPs pour certifier l'origine des matériels de reboisement

De nombreux marqueurs SNPs polymorphes ont été identifiés au sein des populations d'amélioration. De nouveaux multiplex (lots de SNPs) peuvent être développés pour diverses applications concrètes, et notamment pour certifier l'origine des graines utilisées en reboisement. Par exemple, les marqueurs SNPs peuvent permettre de vérifier les provenances des matériels forestiers de reproduction ou de certifier des variétés (vérification de l'origine, de la qualité, de la diversité).

#### ➤ Vérification des provenances des graines utilisées pour le reboisement

Plomion *et al.* (2015) ont mis en évidence un ensemble de marqueurs SNPs très différenciés entre la provenance landaise et des provenances ibériques. Afin de mettre au point une analyse de routine pour discriminer les provenances de pin maritime, un sous-ensemble de 33 SNPs a été sélectionné sur la base de leur niveau de différenciation entre provenances. Ce test a été validé en conditions réelles et permet de certifier la provenance de peuplements candidats au reboisement du massif aquitain. Cette certification présente un réel intérêt car les provenances ibériques (du Portugal et d'une grande partie de l'Espagne) sont proscrites par arrêté ministériel depuis 1986 pour la reconstitution des peuplements en Aquitaine car elles sont inadaptées aux froids hivernaux. En effet, en janvier 1985, le massif landais a connu une vague de froid très intense pendant quelques jours (entre -15°C et -22°C selon les sites) qui a dévasté une surface importante des parcelles boisées (60 000 ha touchés). Après analyse (à l'aide d'un test basé sur la composition terpénique des tissus corticaux), il a été montré que 95 % des peuplements atteints étaient d'origine ibérique (GIS PMF 2002 ; Ribeiro *et al.* 2002).

Aujourd'hui, il reste obligatoire de vérifier la provenance des peuplements porte-graines en vue de leur classement (prérequis nécessaire pour la commercialisation des graines). Ce test basé sur les SNPs remplace maintenant le test terpénique.

#### ➤ Vers la certification des variétés améliorées

Les SNPs permettraient aussi de certifier l'origine des variétés améliorées commercialisées. Par exemple, on pourrait vérifier que les plants vendus proviennent bien d'un verger à graines produisant la variété en question. Dans ce cas, une reconstitution de pedigree avec un jeu de SNPs proche de celui développé dans cette thèse serait une méthode pertinente (le plant provient du verger à graines si l'on retrouve au moins sa mère). On pourrait aussi quantifier la contribution de chaque constituant du verger ainsi que le taux de pollution pollinique.

Enfin, d'autres puces SNPs permettraient de suivre des croisements inter-provenances (combinaisons Landes x Corse, Landes x Maroc, Landes x Portugal) qui sont très étudiés au sein du GIS PMF dans le cadre de la création de nouvelles variétés.



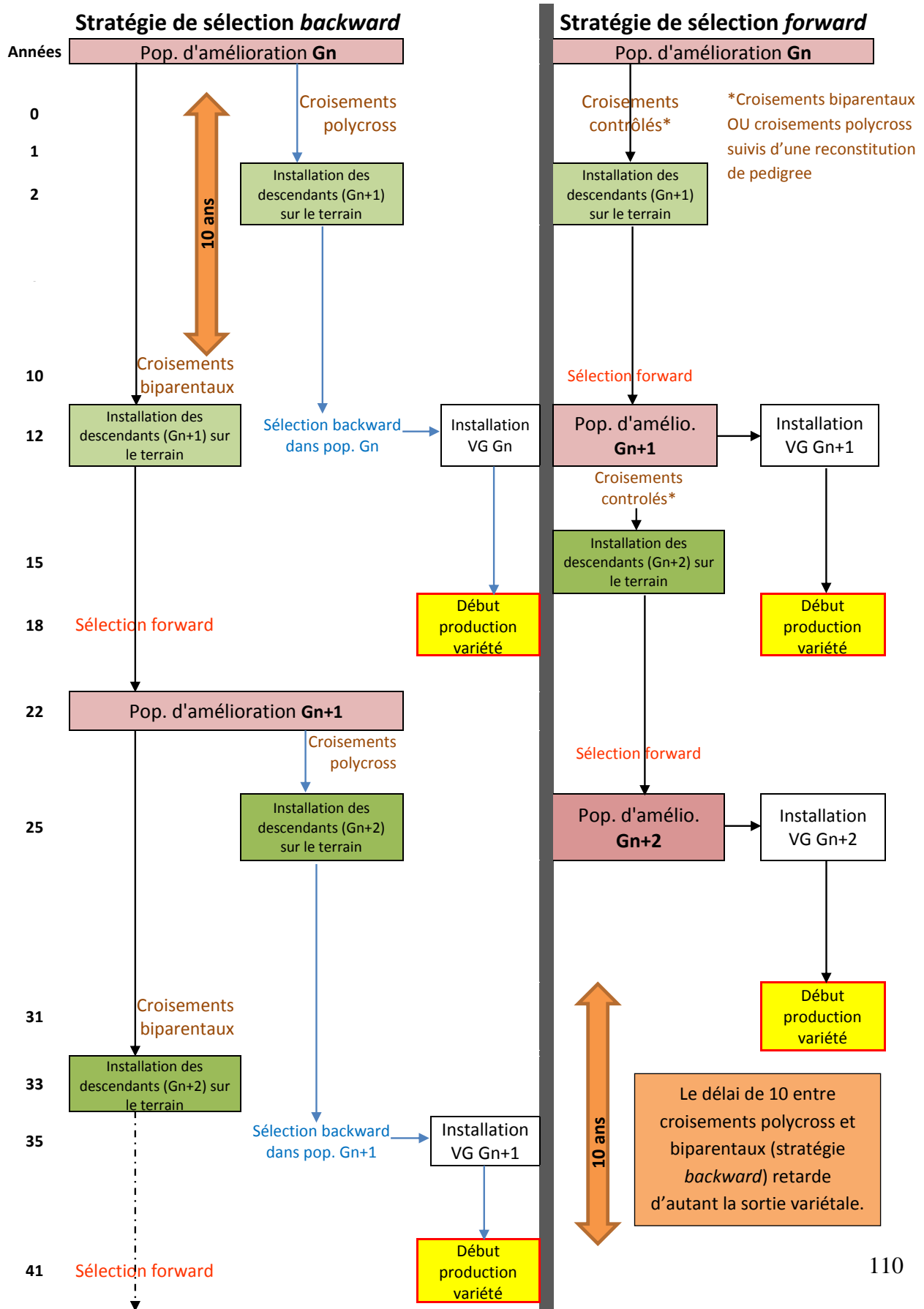
## 6.2 Vers la valorisation des tests de descendance polycross existants ?

Entre 1994 et 2003, près de 100 ha de tests de descendance polycross ont été installés pour tester environ 1000 individus G1. Les résultats présentés dans la *partie 5* montrent qu'il est possible de sélectionner les meilleurs descendants G2 dans ces tests, après génotypage et reconstitution de pedigree, pour constituer une population de production d'une variété améliorée. Notre travail a été réalisé dans un seul test, mais il pourrait être poursuivi dans les autres tests de descendance encore disponibles, dans le but de sélectionner systématiquement de très bons individus G2 pour les intégrer au programme d'amélioration. Les résultats de cette thèse (cf. *partie 5*) montrent qu'un verger issu de la sélection *forward* (d'individus G2 du test de descendance) produit une variété améliorée de gain génétique très légèrement supérieur à une variété produite par un verger issu d'une sélection *backward* (d'individus G1). Cela est notamment dû au fait que la composition des polymix n'est pas optimale : ils n'ont pas été conçus dans une optique de sélection *forward*. Ainsi, au premier abord, il paraît peu intéressant de s'orienter vers la sélection systématique des meilleurs individus des tests existants dans le seul but d'obtenir une variété de gain génétique supérieur. Cependant, ces tests de descendance contiennent de très bons individus G2 qu'il serait dommage de ne pas sélectionner, notamment en terme de diversité. En effet, le programme « *Pinaster* » prévoit un renouvellement variétal plus fréquent et l'installation de nouveaux vergers en continu. La sélection *forward* des individus G2 permettrait donc de diversifier les prochaines sorties variétales en apportant plus de variabilité génétique à gain espéré équivalent. On a d'ailleurs montré que le mélange d'individus G1 (issus de la sélection *backward*) et G2 (issus d'une stratégie de sélection *forward*) dans un même verger à graines constitue une solution intéressante en termes de gain génétique et de diversité.

De plus, on pourrait aussi sélectionner certains individus G2 au sein des tests de descendance polycross pour compléter la population d'amélioration G2 actuelle (issue de croisements biparentaux entre G1). Cela apporterait de nouvelles combinaisons intéressantes de gènes valorisées dans la future population d'amélioration.

**Figure 27 : Comparaison des stratégies de sélection backward et forward dans le cadre du programme d'amélioration du pin maritime.**

Les années sont indiquées à titre d'exemple. On considère que l'évaluation génétique a lieu à 10 ans, et que l'installation d'un verger à graines (VG) est effectuée la même année que celle de la sélection. La production de la variété améliorée débute 6 ans après l'installation du verger. Dans le cadre de la sélection backward, le délai entre les croisements polycross et biparentaux est estimé à 10 ans (ce qui augmente la durée du cycle de sélection de 10 ans).



### 6.3 Vers la mise en place d'une stratégie de sélection *forward* dans le programme d'amélioration

Une stratégie de sélection *forward* consiste à sélectionner les meilleurs individus au sein de la descendance. Afin d'assurer le suivi du pedigree au cours des générations, la sélection *forward* est habituellement effectuée dans la descendance issue de croisements contrôlés biparentaux (les deux parents sont connus). Cependant, les croisements biparentaux présentent plusieurs inconvénients, tant au point de vue logistique que du faible nombre de combinaisons possibles en raison du nombre limité de croisements réalisables chaque année. L'utilisation des marqueurs moléculaires pour reconstituer les pedigrees permettrait de simplifier ces étapes de croisements contrôlés biparentaux. Les croisements polycross sont plus simples à effectuer et à gérer que les croisements biparentaux, et ils permettent de maximiser le nombre de croisements. On pourrait donc remplacer la majorité des croisements biparentaux par des croisements polycross associés à une reconstitution de pedigree, et sélectionner en *forward* les meilleurs descendants (avec un pedigree complet). Ces individus pourraient constituer, d'une part, la population d'amélioration de la génération suivante, et d'autre part, la population de production (géniteurs d'une variété améliorée). Cette stratégie, appelée *polymix breeding* (proposée par Lambeth *et al.* (2001)) permet de réduire la durée du cycle de sélection et par conséquent de réduire aussi le délai entre le choix d'un critère de sélection et la production de la variété améliorée pour ce critère. En effet, dans une stratégie de sélection *backward*, on doit réaliser des croisements polycross afin de créer les tests de descendance (permettant d'évaluer les valeurs génétiques maternelles) mais aussi des croisements biparentaux pour générer la génération suivante de la population d'amélioration. Pour des raisons logistiques, les croisements biparentaux sont réalisés quelques années après les croisements polycross. Par exemple, dans le cas de l'évaluation des individus G1, les croisements biparentaux ont été réalisés environ 10 ans après le début des croisements polycross (qui se sont étalés sur plusieurs années). A noter que ce délai peut être très variable, et dépend du nombre de croisements à réaliser. Dans une stratégie de sélection *forward* de type *polymix breeding*, la génération suivante de la population d'amélioration est issue des croisements polycross associés à une reconstitution de pedigree (Figure 27). Ainsi, le cycle de sélection est plus court (une seule phase de croisements contrôlés au lieu de deux phases distinctes en sélection classique) et les sorties variétales sont plus fréquentes. Si on reprend l'exemple de la Figure 27 (avec un délai de 10 ans entre croisements polycross et croisements biparentaux), on aurait une sortie variétale tous les 12 ans avec une stratégie de sélection *forward*, contre tous les 22 ans avec une stratégie de sélection *backward*. Ce renouvellement variétal plus fréquent facilite l'adaptation des variétés aux évolutions environnementales et économiques, mais permet également d'accroître le gain génétique par une intensité de sélection plus forte sur les géniteurs des vergers à graines. En effet, la plus faible diversité génétique à l'intérieur de chaque variété serait compensée, au niveau du massif, par un plus grand nombre de variétés différentes (Bouffier et Pastuszka 2014).

De plus, cette stratégie de sélection *forward* de type *polymix breeding* permet de corriger les erreurs d'identités et de pedigree. Cela permet d'améliorer la précision des prédictions des valeurs génétiques et de gagner en efficacité. L'utilisation des marqueurs moléculaires pour la reconstitution de pedigree présente donc un double intérêt : on simplifie les étapes de croisements, et le suivi du pedigree est plus fiable.



En sélection *forward*, la valeur génétique d'un descendant évalué est estimée à partir de ses propres performances mais aussi celles de l'ensemble de ses apparentés (ascendants et collatéraux). Cependant, son estimation est moins précise que lorsqu'elle est estimée sur descendance, ce qui peut constituer un obstacle à sa mise en place. Dans le cadre du programme d'amélioration du pin maritime, les performances de plus de 450 000 arbres sont enregistrées dans la base de données Dataplan (Bouffier et Pastuszka 2014) alimentant le système d'évaluation génétique TreePlan (moteur d'analyse génétique multigénérationnel optimisé pour les programmes d'amélioration des arbres forestiers). Ainsi, l'analyse combinée d'un grand nombre de données permet une estimation assez précise des valeurs génétiques des descendants candidats à la sélection *forward*.

Avant de mettre en place ce type de stratégie au sein du programme d'amélioration du pin maritime, des études de simulation sont nécessaires pour déterminer le nombre de polymix à utiliser et leur composition (nombre de pollens, apparentements intra et inter polymix) et pour optimiser les plans de croisements (groupes de clones à croiser avec un ou plusieurs polymix), afin de maximiser les gains génétiques tout en contrôlant l'apparentement au fil des générations. Il faudrait aussi affiner les méthodes de présélection (choix des individus à génotyper) et de sélection par simulation, par exemple en utilisant le logiciel POPSIM<sup>TM</sup> (Mullin *et al.* 2010). Ce logiciel permet, par des méthodes de simulation stochastiques, de tester différents plans de croisements, différentes méthodes de sélection et d'estimer pour chaque scénario des gains génétiques et la taille efficace de la population, et cela sur plusieurs générations.

#### 6.4 Vers la sélection génomique

Dans ce travail de thèse, les marqueurs moléculaires ont été utilisés pour reconstituer les pedigrees, mais l'évolution des techniques de génotypage à haut débit permet d'envisager de nouvelles méthodes de sélection intégrant un plus grand nombre de marqueurs. En effet, il est maintenant possible de caractériser des individus pour un très grand nombre de marqueurs (quelques milliers de SNPs), rapidement, et à un coût plus accessible (et qui devrait encore décroître ces prochaines années). L'information obtenue par un plus grand nombre de marqueurs permet d'estimer plus précisément la proportion d'allèles partagés par deux individus de la même famille : ainsi, de manière analogue à la matrice de parenté **A** (calculée à partir de l'information du pedigree) on peut calculer une matrice de parenté génomique **G**. Les éléments de cette matrice **G** correspondent aux relations génomiques entre les individus estimées à partir des marqueurs. L'estimation des apparentements génomiques est plus précise que celle des apparentements généalogiques (sous condition d'une quantité suffisante de marqueurs). En effet, la généalogie peut être incomplète ou incorrecte, et la matrice **A** (parenté espérée) ne prend pas en compte les portions de génome réellement partagées entre individus apparentés, alors que les relations génomiques précisent la ségrégation mendélienne et décrivent la parenté « réelle » observée à travers les marqueurs (Isik 2014). Ces relations génomiques permettent d'estimer les GEBV (genomic estimated breeding value) avec le modèle G-BLUP.

D'autres modèles d'évaluation génomique, basés sur les statistiques bayésiennes, prédisent les effets des marqueurs : l'information génomique d'un ensemble d'individus génotypés (variables explicatives) est analysée pour expliquer des performances observées

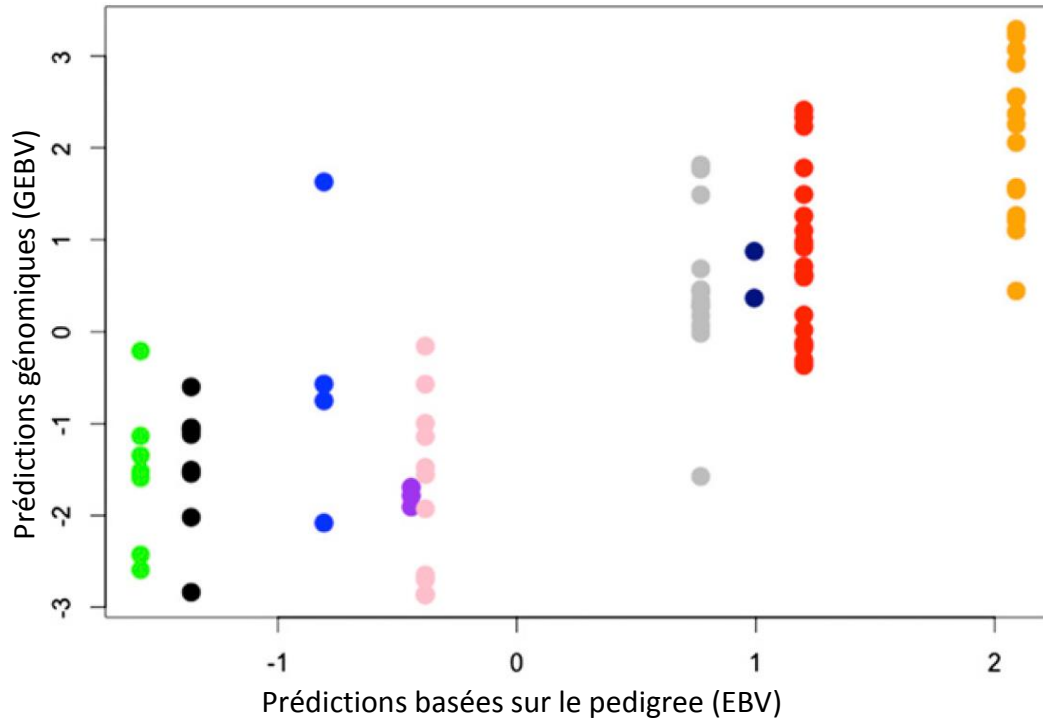
(variables à expliquer) afin de construire le meilleur modèle prédictif de la valeur génomique (GEBV), qui est la somme des effets des marqueurs (Meuwissen *et al.* 2001).

Des études de simulation ont montré l'intérêt potentiel que représente la sélection génomique en amélioration forestière (Denis et Bouvet 2013 ; Grattapaglia et Resende 2011 ; Iwata *et al.* 2011), notamment grâce à la possibilité d'une sélection plus précoce qui permet d'accélérer les cycles de sélection (et donc d'augmenter le gain par unité de temps). Quelques études empiriques (Beaulieu *et al.* 2014 ; Isik *et al.* 2016 ; Resende *et al.* 2012a ; Resende *et al.* 2012b ; Zapata-Valenzuela *et al.* 2012 ; Zapata-Valenzuela *et al.* 2013) montrent que les précisions des modèles de prédiction génomique sont équivalentes voire inférieures aux précisions des prédictions des valeurs génétiques (EBV, estimated breeding value) estimées avec l'information du pedigree et des données phénotypiques. Par exemple, selon une étude sur *Pinus taeda* de Zapata-Valenzuela *et al.* (2012) basée sur 3406 SNPs, les précisions des prédictions génomiques varient entre 0,30 et 0,68 pour la hauteur et le volume. Resende *et al.* (2012a) ont étudié la précision de la sélection génomique sur *Eucalyptus* pour des caractères de qualité du bois et de croissance : elle varie entre 0,55 et 0,88 (selon les caractères étudiés et la population étudiée). Cette précision est comparable ou inférieure à la précision des estimations des EBV basées sur le phénotype et le pedigree. Sur le pin maritime, une première étude (Isik *et al.* 2016), basée sur le génotypage de 661 individus (générations G0 et G1) avec 2500 SNPs, montre que les précisions de la prédiction génomique varient entre 0,39 et 0,55 selon la méthode de validation utilisée et le caractère étudié (hauteur, circonférence, rectitude du tronc). Une autre étude est actuellement en cours sur trois générations (G0, G1, G2) pour évaluer l'aptitude des modèles de sélection génomique à prédire les performances des descendants. En effet, la validation des modèles de sélection génomique entre générations est indispensable dans une optique de sélection et constitue une première chez les arbres forestiers avec des données réelles (Bartholomé *et al.* 2015). Trois types de modèles (un basé sur le pedigree et deux sur les marqueurs) ont été comparés. 818 individus (710 G2 et tous leurs ascendants, soit 62 G1 et 46 G0) ont été génotypés avec 4300 SNPs informatifs. En moyenne, le modèle basé sur le pedigree a obtenu des précisions équivalentes à celles des modèles de prédiction génomique. Pour la validation sur descendance (population de calibration constituée par les individus G0 et G1, et population de validation constituée par les 710 individus G2), une précision de prédiction des GEBV de 0,70 à 0,81 a été obtenue.

Compte tenu du faible nombre de marqueurs (2500 à 7000 SNPs) utilisés dans toutes ces études, la précision des modèles de prédiction génomique repose surtout sur le fait que les marqueurs précisent les apparentements réels entre parents et descendants en révélant la ségrégation mendélienne des allèles, et non pas sur l'exploitation des déséquilibres de liaison « historiques » entre marqueurs SNPs et QTL (Habier *et al.* 2013). De plus, les tailles des populations d'étude sont réduites (ce qui favorise le déséquilibre de liaison). Les modèles de prédiction génomique ne sont donc pas transposables directement à des populations d'amélioration à forte variabilité génétique et à faible déséquilibre de liaison.

Malgré les obstacles à la mise en place de la sélection génomique proprement dite au sein des programmes d'amélioration génétique des arbres forestiers, plusieurs applications sont envisagées pour le pin maritime.

**Figure 28** (issue de Zapata-Valenzuela et al. (2013)) : Prédiction des valeurs génétiques de plein-frères issus de 9 croisements, basées sur le pedigree (EBV, axe x) ou sur les apparentements génomiques (GEBV, axe y). Dans chaque famille, sans connaissance du phénotype, les valeurs génétiques des plein-frères, estimées à l'aide de la matrice de parenté  $A$  (EBV), sont identiques et égales à la moyenne de leurs deux parents. Les données génomiques permettent d'estimer les GEBV et de distinguer les plein-frères entre-eux.



Par exemple, le gain de précision des prédictions obtenues avec les données génomiques nous permettrait de phénotyper les candidats à la sélection plus précocement (par exemple en mesurant la croissance à 8 ans au lieu de 12 ans), et ainsi avancer de 4 ans l'âge de la sélection. On pourrait aussi sélectionner sur des caractères complexes et coûteux à mesurer, pour lesquels le phénotypage sur un très grand nombre d'individus est économiquement inenvisageable (ex: résistance aux stress abiotique et biotique). Dans ce cas, une partie de la population serait génotypée et phénotypée (pour établir le modèle de prédiction génomique) et on pourrait prédire la valeur génétique GEBV du reste de la population à sélectionner à partir des génotypes uniquement. Les données génomiques permettraient également de sélectionner les meilleurs individus au sein de familles de plein-frères (Figure 28), grâce à la prise en compte de l'effet mendélien.

A l'heure actuelle, il est trop coûteux de génotyper l'ensemble de la population avec un grand nombre de marqueurs. La matrice de parenté  $\mathbf{A}$  peut alors être remplacée par une matrice hybride  $\mathbf{H}$  qui prend en compte à la fois l'information des relations de parenté classiques ( $\mathbf{A}$ ) et l'information génomique ( $\mathbf{G}$ ) (Legarra *et al.* 2009 ; Misztal *et al.* 2009). Ainsi on intègre simultanément les individus non génotypés (notés 1) et les individus génotypés (notés 2). La matrice de parenté  $\mathbf{A}$  s'écrit alors :

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

La matrice  $\mathbf{A}_{11}$  contient les relations de parenté entre individus non génotypés, les matrices  $\mathbf{A}_{12}$  et  $\mathbf{A}_{21}$  les relations de parenté entre individus non génotypés et génotypés, et la matrice  $\mathbf{A}_{22}$  les relations de parenté entre individus génotypés (*i.e.* la matrice  $\mathbf{G}$ ). La matrice  $\mathbf{H}$  peut donc s'écrire :

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{G} \end{bmatrix} = \mathbf{A} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G} - \mathbf{A}_{22} \end{bmatrix}$$

Pour simplifier les calculs, on utilise la matrice inverse  $\mathbf{H}^{-1}$  (Aguilar *et al.* 2010) :

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Cette matrice hybride permet donc de valoriser une grande partie des données du programme d'amélioration (performances mesurées sur plusieurs générations, pedigrees enregistrés), et de compléter avec des données de génomique (acquises progressivement) sur une partie de la population. Cette approche permettrait d'améliorer la précision des prédictions des valeurs génétiques des candidats à la sélection.

Dans ce contexte, la reconstitution de pedigree garde un intérêt car elle peut constituer une aide précieuse pour améliorer la précision des modèles de prédiction génomique. De plus, elle permet de supprimer les erreurs des bases de données sur lesquelles vont reposer la construction des modèles de prédiction génomique en vérifiant les identités génétiques et les pedigrees des individus constituant la population de calibration du modèle.

Enfin, même si la sélection génomique proprement dite est encore loin d'être appliquée concrètement dans les programmes d'amélioration des essences forestières, les apparentements génomiques (estimés sur un grand nombre de marqueurs) pourraient être utilisés à moyen terme pour améliorer l'efficacité des programmes de sélection.



## 7. Conclusion générale

Ce travail de thèse avait pour objectif d'étudier l'utilisation des marqueurs moléculaires comme un nouvel outil pour optimiser les stratégies d'amélioration du pin maritime. Tout d'abord, j'ai développé un jeu de marqueurs (63 SNPs) pour reconstituer *a posteriori* les pedigrees dans un test de descendance polycross. Dans un second temps, j'ai vérifié les hypothèses sur lesquelles repose la sélection *backward* (*i.e.* familles de demi-frères et égale contribution des pères des polymix). Puis j'ai démontré que les estimations des paramètres génétiques et des valeurs génétiques maternelles, basées sur l'information d'un pedigree partiel ou complet, diffèrent peu. Enfin, j'ai montré qu'il est possible de valoriser les tests de descendance existants en y sélectionnant en *forward* les meilleurs descendants après reconstitution de leur pedigree.

Cette thèse ouvre de nouvelles perspectives vers la mise en place d'une stratégie de sélection *forward* de type *polymix breeding*, qui permet notamment de regrouper en une seule étape les phases de croisements biparentaux et d'évaluation génétique à l'aide d'un ou plusieurs croisements polycross suivis d'une reconstitution de pedigree. Cette stratégie innovante permet de simplifier la phase des croisements contrôlés, de maximiser le nombre de croisements réalisés et d'accélérer les cycles de sélection.

Enfin, cette étude pourrait aussi être élargie à d'autres espèces forestières (*Douglas-fir*, *Pinus taeda*, *Pinus sylvestris*, *Cryptomeria japonica*, *etc.*) afin d'optimiser les stratégies d'amélioration génétique et d'adapter au mieux les variétés améliorées aux changements environnementaux et économiques à venir.



# Bibliographie

- Adams WT, Neale DB, Loopstra CA (1988) **Verifying controlled crosses in conifer tree-improvement programs**. *Silvae Genet* 37:147-152
- Agreste (2015) **Agreste Infos rapides. Bois et dérivés**. Octobre 2015. N°2/2. <http://agreste.agriculture.gouv.fr/IMG/pdf/conjinforap201510bois.pdf>
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) **Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score**. *J Dairy Sci* 93:743-752. doi:10.3168/jds.2009-2730
- Anderson EC, Garza JC (2006) **The power of single-nucleotide polymorphisms for large-scale parentage inference**. *Genetics* 172:2567-2582. doi:10.1534/genetics.105.048074
- Askew GR (1988) **Estimation of gamete pool compositions in clonal seed orchards**. *Silvae Genet* 37:227-232
- Banos G, Wiggans GR, Powell RL (2001) **Impact of paternity errors in cow identification on genetic evaluations and international comparisons**. *J Dairy Sci* 84:2523-2529
- Baradat P (2000) **DIOGENE. Un logiciel intégré pour la génétique et l'amélioration des plantes**. [http://amap-collaboratif.cirad.fr/pages\\_logiciels/uploads/Documents/baradat/notidio.pdf](http://amap-collaboratif.cirad.fr/pages_logiciels/uploads/Documents/baradat/notidio.pdf).
- Bartholomé J, Vidal M, Van Heerwaarden J, Isik F, Plomion C, Bouffier L (2015) **Genomic selection in a multi-generation conifer breeding population: maritime pine as a case study**. In: IUFRO Tree Biotechnology Conference, Florence, Italy, June 2015.
- Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J (2014) **Genomic selection accuracies within and between environments and small breeding groups in white spruce**. *BMC Genomics* 15:16. doi:10.1186/1471-2164-15-1048
- Bell JC, Powell M, Devey ME, Moran GF (2004) **DNA profiling, pedigree lineage analysis and monitoring in the Australian breeding program of radiata pine**. *Silvae Genet* 53:130-134
- Bidanel J, Ducrocq V, Ollivier L (1990) **Les nouvelles méthodes d'évaluation des reproducteurs : intérêt et applications du modèle individuel chez le porc**. *Journées Rech. Porcine en France* 22:1-10.
- Boichard D, Guillaume F, Baur A, Croiseau P, Rossignol MN, Boscher MY, Druet T, Genestout L, Colleau JJ, Journaux L, Ducrocq V, Fritz S (2012) **Genomic selection in French dairy cattle**. *Anim Prod Sci* 52:115-120. doi:10.1071/an11119
- Bonga JM (2015) **A comparative evaluation of the application of somatic embryogenesis, rooting of cuttings, and organogenesis of conifers**. *Can J For Res* 45:379-383. doi:dx.doi.org/10.1139/cjfr-2014-0360
- Bouffier L, Pastuszka P (2014) **Choix du matériel végétal : exemple du massif landais**. *Innovations Agronomiques* 41:31-41
- Bouffier L, Raffin A, Kremer A (2008) **Evolution of genetic variation for selected traits in successive breeding populations of maritime pine**. *Heredity* 101:156-165. doi:10.1038/hdy.2008.41
- Bouffier L, Raffin A, Kremer A (2009) **La gestion de la variabilité génétique dans le programme d'amélioration du pin maritime**. *Revue Forestière Française* 61:369-388. doi:10.4267/2042/30547
- Bramlett D, Burris L (1995) **Topworking young scions into reproductively-mature loblolly pine**. Paper presented at the Proceedings of the 23rd Southern Forest Tree Improvement Conference, Asheville, NC,
- Bramlett DL (1997) **Genetic gain from mass controlled pollination and topworking**. *Journal of Forestry* 95:15-19



- Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alia R, Vendramin GG (2007) **Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers.** Mol Ecol 16:2137-2153. doi:10.1111/j.1365-294X.2007.03275.x
- Burban C, Petit RJ (2003) **Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance.** Mol Ecol 12:1487-1495. doi:10.1046/j.1365-294X.2003.01817.x
- Burczyk J, Nikkanen T, Lewandowski A (1997) **Evidence of an unbalanced mating pattern in a seed orchard composed of two larch species.** Silvae Genet 46:176-181
- Burdon RD, Kumar S (2004) **Forwards versus backwards selection: trade-offs between expected genetic gain and risk avoidance.** N Z J Forest Sci 34:3-21
- Bush D, Kain D, Matheson C, Kanowski P (2011) **Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters-an example using *Eucalyptus cladocalyx*.** Tree Genet Genomes 7:23-35. doi:10.1007/s11295-010-0312-z
- Carle J, Holmgren P (2008) **Wood from planted forests: a global outlook 2005-2030.** Forest Products Society 58:6-18
- Chagne D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C (2003) **Comparative genome and QTL mapping between maritime and loblolly pines.** Mol Breed 12:185-195. doi:10.1023/a:1026318327911
- Chancerel E, Lamy JB, Lesur I, Noirot C, Klopp C, Ehrenmann F, Boury C, Le Provost G, Label P, Lalanne C, Leger V, Salin F, Gion JM, Plomion C (2013) **High-density linkage mapping in a pine tree reveals a genomic region associated with inbreeding depression and provides clues to the extent and distribution of meiotic recombination.** BMC Biol 11:19. doi:10.1186/1741-7007-11-50
- Codesido V, Zas R, Fernandez-Lopez J (2012) **Juvenile-mature genetic correlations in *Pinus radiata* D. Don. under different nutrient x water regimes in Spain.** European Journal of Forest Research 131:297-305. doi:10.1007/s10342-011-0501-9
- Cornelius J (1994) **Heritabilities and additive genetic coefficients of variation in forest trees.** Can J For Res 24:372-379. doi:10.1139/x94-050
- Denis M, Bouvet JM (2013) **Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding.** Tree Genet Genomes 9:37-51. doi:10.1007/s11295-012-0528-1
- Desprez-Loustau M, Marcais B, Nageleisen L, Piou D, Vannini A (2006) **Interactive effects of drought and pathogens in forest trees.** Ann For Sci 63:597-612
- Devey ME, Carson SD, Nolan MF, Matheson AC, Riini CT, Hohepa J (2004) **QTL associations for density and diameter in *Pinus radiata* and the potential for marker-aided selection.** Theor Appl Genet 108:516-524. doi:10.1007/s00122-003-1446-2
- Doerksen TK, Herbinger CM (2008) **Male reproductive success and pedigree error in red spruce open-pollinated and polycross mating systems.** Can J For Res-Rev Can Rech For 38:1742-1749. doi:10.1139/x08-025
- Doerksen TK, Herbinger CM (2010) **Impact of reconstructed pedigrees on progeny-test breeding values in red spruce.** Tree Genet Genomes 6:591-600. doi:10.1007/s11295-010-0274-1
- Durel CE, Bertin P, Kremer A (1996) **Relationship between inbreeding depression and inbreeding coefficient in maritime pine (*Pinus pinaster*).** Theor Appl Genet 92:347-356
- Edwards DGW, ElKassaby YA (1996) **The biology and management of coniferous forest seeds: genetic perspectives.** For Chron 72:481-484
- El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klapste J, Lstiburek M (2011) **Breeding without Breeding: is a complete pedigree necessary for efficient breeding?** PLoS One 6:11. doi:10.1371/journal.pone.0025737

- El-Kassaby YA, Funda T, Lai BSK (2010) **Female reproductive success variation in a *Pseudotsuga menziesii* seed orchard as revealed by pedigree reconstruction from a bulk seed collection.** J Hered 101:164-168. doi:10.1093/jhered/esp126
- El-Kassaby YA, Funda T, Liewlaksaneeyanawin C (2015) **Increasing Breeding without Breeding (BwB) efficiency: full- vs. partial-pedigree reconstruction in Lodgepole Pine.** SOJ Genet Sci 2(1):1-6
- El-Kassaby YA, Lstiburek M (2009) **Breeding without breeding.** Genet Res 91:111-120. doi:10.1017/s001667230900007x
- Evans J (2009) **Planted forests: uses, impacts and sustainability.** CABI and FAO.
- Falconer DS, Mackay TFC (1996) **Introduction to quantitative genetics.** vol Ed. 4. Longman Group Limited, Harlow, UK
- FAO (2010) **Évaluation des ressources forestières mondiales 2010: Rapport principal.** Étude FAO-Forêts 163.Rome.
- FAO (2014) **Situation des forêts du monde,** Rome.
- FAO (2015) **Evaluation des ressources forestières mondiales 2015: comment les forêts de la planète changent-elles?** Rome.
- Fritz S, Colleau JJ, Druet T, Boscher MY, Rossignol MN, Malafosse A, Boichard D (2003) **Mise en place d'une Sélection Assistée par Marqueurs dans les trois principales races bovines laitières françaises.** Renc Rech Ruminants 10:53-56
- Funda T, El-Kassaby YA (2012) **Seed orchard genetics.** CAB Reviews 7:1-23. doi:10.1079/pavsnr20127013
- Funda T, Liewlaksaneeyanawin C, El-Kassaby YA (2014) **Determination of paternal and maternal parentage in lodgepole pine seed: full versus partial pedigree reconstruction.** Can J For Res-Rev Can Rech For 44:1122-1127. doi:10.1139/cjfr-2014-0145
- Funda T, Lstiburek M, Lachout P, Klapste J, El-Kassaby YA (2009) **Optimization of combined genetic gain and diversity for collection and deployment of seed orchard crops.** Tree Genet Genomes 5:583-593. doi:10.1007/s11295-009-0211-3
- Funda T, Wennström U, Almqvist C, Torimaru T, Gull B-A, Wang X-R (2015) **Low rates of pollen contamination in a Scots pine seed orchard in Sweden: the exception or the norm?** Scand J Forest Res 30:573-586. doi:10.1080/02827581.2015.1036306
- Gabriel S, Ziaugra L, Tabbaa D (2009) **SNP genotyping using the Sequenom MassARRAY iPLEX Platform.** Current Protocols in Human Genetics, Chapter 2:Unit 212. doi:10.1002/0471142905.hg0212s60
- Gauzere J, Oddou-Muratorio S, Pichot C, Lefevre F, Klein E (2013) **Biases in quantitative genetic analyses using open-pollinated progeny tests from natural tree populations.** Acta Bot Gall 160:225-236. doi:10.1080/12538078.2013.822827
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) **ASReml User Guide Release 3.0.** VSN International Ltd, Hemel Hempstead, HP1 1ES, UK
- GIS PMF (2002) Alazard P, Canteloup D, Crémière L, Daubet A, Lesgourgues Y, Merzeau D, Pastuska P, Raffin A. **Le progrès génétique en forêt.** <http://www.mediaforest.net/francais/sciences/gis.htm>
- GIS PMF (2014) GIS Groupe Pin maritime du futur. **Les cahiers de la reconstitution n°4: matériel végétal de reboisement.** <http://www.onf.fr/outils/medias/20130708-143100-661300/++files++/4>.
- Gomory D, Bruchanik R, Longauer R (2003) **Fertility variation and flowering asynchrony in *Pinus sylvestris*: consequences for the genetic structure of progeny in seed orchards.** For Ecol Manage 174:117-126. doi:10.1016/s0378-1127(02)00031-2
- Grattapaglia D, Diener PSD, dos Santos GA (2014) **Performance of microsatellites for parentage assignment following mass controlled pollination in a clonal seed orchard of loblolly pine (*Pinus taeda* L.).** Tree Genet Genomes 10:1631-1643. doi:10.1007/s11295-014-0784-3

- Grattapaglia D, Resende MDV (2011) **Genomic selection in forest tree breeding**. *Tree Genet Genomes* 7:241-255. doi:10.1007/s11295-010-0328-4
- Grattapaglia D, Ribeiro VJ, Rezende G (2004) **Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus***. *Theor Appl Genet* 109:192-199. doi:10.1007/s00122-004-1617-9
- Guimarães E, Ruane J, Scherf BD, Sonnino A, J.D. D (2007) **Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish**. Paper presented at the FAO, Rome,
- Habier D, Fernando RL, Garrick DJ (2013) **Genomic BLUP decoded: A look into the black box of genomic prediction**. *Genetics* 194:597-+. doi:10.1534/genetics.113.152207
- Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) **Factors influencing levels of genetic diversity in woody plant species**. *New For* 6:95-124. doi:10.1007/bf00120641
- Hansen OK (2008) **Mating patterns, genetic composition and diversity levels in two seed orchards with few clones - Impact on planting crop**. *For Ecol Manage* 256:1167-1177. doi:10.1016/j.foreco.2008.06.032
- Hansen OK, McKinney LV (2010) **Establishment of a quasi-field trial in *Abies nordmanniana*- test of a new approach to forest tree breeding**. *Tree Genet Genomes* 6:345-355. doi:10.1007/s11295-009-0253-6
- Hansen OK, Nielsen UB (2010) **Microsatellites used to establish full pedigree in a half-sib trial and correlation between number of male strobili and paternal success**. *Ann For Sci* 67:10. doi:10.1051/forest/2010028
- Hayes B, Goddard M (2010) **Genome-wide association and genomic selection in animal breeding**. *Genome* 53:876-883. doi:10.1139/g10-076
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) **Genomic selection in dairy cattle: progress and challenges**. *J Dairy Sci* 92:433-443. doi:10.3168/jds.2008-1646
- Henderson CR (1950) **Estimation of genetic parameters**. *Ann Math Statist* 21:309-310
- Henderson CR (1963) **Selection index and the expected genetic advance**. In: Hanson WD, Robinson HF (Eds). *Statistical genetics and plant breeding*, pp. 141-163.
- Henderson CR (1975) **Best linear unbiased estimation and prediction under a selection model**. *Biometrics* 31:423-447
- Heslot N, Jannink JL, Sorrells ME (2015) **Perspectives for genomic selection applications and research in plants**. *Crop Sci* 55:1-12. doi:10.2135/cropsci2014.03.0249
- Hosius B, Bergmann F, Konnert M, Henkel W (2000) **A concept for seed orchards based on isoenzyme gene markers**. *For Ecol Manage* 131:143-152. doi:10.1016/s0378-1127(99)00209-1
- IFN (2009) **Inventaire forestier national n°21. Tempête Klaus du 24 janvier 2009**. [http://inventaire-forestier.ign.fr/spip/IMG/pdf/IF21\\_internet.pdf](http://inventaire-forestier.ign.fr/spip/IMG/pdf/IF21_internet.pdf).
- Illy G (1966) **Recherches sur l'amélioration génétique du pin maritime**. *Ann Sci forest* 23:765-948. doi:10.1051/forest/19660401
- Illy G, Sopena J (1963) **La dispersion du pollen de Pin maritime**. *Revue Forestière Française* 15:7-18
- Isik F (2014) **Genomic selection in forest tree breeding: the concept and an outlook to the future**. *New For* 45:379-401. doi:10.1007/s11056-014-9422-z
- Isik F, Bartholome J, Farjat A, Chancerel E, Raffin A, Sanchez L, Plomion C, Bouffier L (2016) **Genomic selection in maritime pine**. *Plant Science* 242:108-119. doi:10.1016/j.plantsci.2015.08.006
- Israel C, Weller JI (2000) **Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations**. *J Dairy Sci* 83:181-187
- IUFRO (2015) **Building up the Bio-Economy: the forest sector has to take the lead**. Vienna

- Iwata H, Hayashi T, Tsumura Y (2011) **Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica***. *Tree Genet Genomes* 7:747-758. doi:10.1007/s11295-011-0371-9
- Jaramillo-Correa JP, Rodriguez-Quilon I, Grivet D, Lepoittevin C, Sebastiani F, Heuertz M, Garnier-Gere PH, Alia R, Plomion C, Vendramin GG, Gonzalez-Martinez SC (2015) **Molecular proxies for climate maladaptation in a long-lived tree (*Pinus pinaster Aiton*, Pinaceae)**. *Genetics* 199:793-+. doi:10.1534/genetics.114.173252
- Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN, Adams WT, Neale DB (2001) **Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness**. *Theor Appl Genet* 102:1152-1158. doi:10.1007/s001220000506
- Johnson R, Lipow S (2002) **Compatibility of breeding for increased wood production and long-term sustainability: the genetic variation of seed orchard seed and associated risks**. In: Johnson AC, Haynes RW, Monserud RA (eds) *Congruent Management of Multiple Resources: Proceedings from the Wood Compatibility Initiative Workshop*, vol 563. Usda Forest Service General Technical Report Pacific Northwest. Us Dept Agr, Forest Serv Pacific Nw Research Stn, Portland, pp 169-179
- Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) **A practical guide to methods of parentage analysis**. *Mol Ecol Resour* 10:6-30. doi:10.1111/j.1755-0998.2009.02778.x
- Kalinowski ST, Taper ML, Marshall TC (2007) **Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment**. *Mol Ecol* 16:1099-1106. doi:10.1111/j.1365-294X.2007.03089.x
- Kang KS, Lindgren D (1998) **Fertility variation and its effect on the relatedness of seeds in *Pinus densiflora*, *Pinus thunbergii* and *Pinus koraiensis* clonal seed orchards**. *Silvae Genet* 47:196-201
- Klapste J, Lstiburek M, El-Kassaby YA (2014) **Estimates of genetic parameters and breeding values from western larch open-pollinated families using marker-based relationship**. *Tree Genet Genomes* 10:241-249. doi:10.1007/s11295-013-0673-1
- Korecky J, Klapste J, Lstiburek M, Kobliha J, Nelson CD, El-Kassaby YA (2013) **Comparison of genetic parameters from marker-based relationship, sibship, and combined models in Scots pine multi-site open-pollinated tests**. *Tree Genet Genomes* 9:1227-1235. doi:10.1007/s11295-013-0630-z
- Kremer A (1992) **Predictions of age-age correlations of total height based on serial correlations between height increments in Maritime pine (*Pinus pinaster Ait.*)**. *Theor Appl Genet* 85:152-158
- Kumar S, Gerber S, Richardson TE, Gea L (2007) **Testing for unequal paternal contributions using nuclear and chloroplast SSR markers in polycross families of radiata pine**. *Tree Genet Genomes* 3:207-214. doi:10.1007/s11295-006-0056-y
- Kumar S, Richardson TE (2005) **Inferring relatedness and heritability using molecular markers in radiata pine**. *Mol Breed* 15:55-64. doi:10.1007/s11032-004-2059-4
- Lai BS, Funda T, Liewlaksaneeyanawin C, Klapste J, Van Niejenhuis A, Cook C, Stoehr MU, Woods J, El-Kassaby YA (2010) **Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction**. *Ann For Sci* 67:8. doi:10.1051/forest/2010044
- Lambeth C, Lee BC, O'Malley D, Wheeler N (2001) **Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing**. *Theor Appl Genet* 103:930-943. doi:10.1007/s001220100627
- Lambeth CC (1980) **Juvenile-mature correlations in Pinaceae and implications for early selection**. *For Sci* 26:571-580
- Legarra A, Aguilar I, Misztal I (2009) **A relationship matrix including full pedigree and genomic information**. *J Dairy Sci* 92:4656-4663. doi:10.3168/jds.2009-2061

- Lelu-Walter MA, Thompson D, Harvengt L, Sanchez L, Toribio M, Paques LE (2013) **Somatic embryogenesis in forestry with a focus on Europe: state-of-the-art, benefits, challenges and future direction**. *Tree Genet Genomes* 9:883-899. doi:10.1007/s11295-013-0620-1
- 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 Genet Genomes* 8:113-126. doi:10.1007/s11295-011-0426-y
- Lerceteau E, Plomion C, Andersson B (2000) **AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study**. *Mol Breed* 6:451-458. doi:10.1023/a:1026548716320
- Lindgren D, Danusevicius D, Rosvall O (2009) **Unequal deployment of clones to seed orchards by considering genetic gain, relatedness and gene diversity**. *Forestry* 82:17-28. doi:10.1093/forestry/cpn033
- Lindgren D, Gea LD, Jefferson PA (1997) **Status number for measuring genetic diversity**. *Forest Genetics* 4:69-76
- Lindgren D, Kang K (1997) **Status number - a useful tool for tree breeding**. Research Report of the Forest Genetics Research Institute (Suwon):154-165
- Lindgren D, Mullin TJ (1998) **Relatedness and status number in seed orchard crops**. *Can J For Res-Rev Can Rech For* 28:276-283. doi:10.1139/cjfr-28-2-276
- Lstiburek M, Hodge GR, Lachout P (2015) **Uncovering genetic information from commercial forest plantations-making up for lost time using "Breeding without Breeding"**. *Tree Genet Genomes* 11:12. doi:10.1007/s11295-015-0881-y
- Lstiburek M, Ivankova K, Kadlec J, Kobliha J, Klapste J, El-Kassaby YA (2011) **Breeding without breeding: minimum fingerprinting effort with respect to the effective population size**. *Tree Genet Genomes* 7:1069-1078. doi:10.1007/s11295-011-0395-1
- Machanska E, Bajcar V, Longauer R, Gomory D (2013) **Effective population size estimation in seed orchards: a case study of *Pinus nigra* ARNOLD and *Fraxinus excelsior* L./ *F. angustifolia* VAHL**. *Genetika-Belgrade* 45:575-588. doi:10.2298/genr1302575m
- Mackay J, Dean JFD, Plomion C, Peterson DG, Canovas FM, Pavy N, Ingvarsson PK, Savolainen O, Guevara MA, Fluch S, Vinceti B, Abarca D, Diaz-Sala C, Cervera MT (2012) **Towards decoding the conifer giga-genome**. *Plant MolBiol* 80:555-569. doi:10.1007/s11103-012-9961-7
- Malécot G (1948) **Les mathématiques de l'hérédité**. Masson & Cie, Paris, 64p.
- Mallez S (2014) **Invasion du nématode du pin *Bursaphelenchus xylophilus***. Sciences agricoles. Université Nice Sophia Antipolis. <https://tel.archives-ouvertes.fr/tel-01127300>.
- Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2012) **Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness**. *Heredity* 108:285-291. doi:10.1038/hdy.2011.73
- Mardis ER (2008) **Next-generation DNA sequencing methods**. In: Annual Review of Genomics and Human Genetics, vol 9. Annual Review of Genomics and Human Genetics. Annual Reviews, Palo Alto, pp 387-402. doi:10.1146/annurev.genom.9.081307.164359
- Marguerit E, Bouffier L, Chancerel E, Costa P, Lagane F, Guehl JM, Plomion C, Brendel O (2014) **The genetics of water-use efficiency and its relation to growth in maritime pine**. *J Exp Bot* 65:4757-4768. doi:10.1093/jxb/eru226
- Mariette S, Chagne D, Lezier C, Pastuszka P, Baffin A, Plomion C, Kremer A (2001) **Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers**. *Heredity* 86:469-479. doi:10.1046/j.1365-2540.2001.00852.x
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) **Statistical confidence for likelihood-based paternity inference in natural populations**. *Mol Ecol* 7:639-655. doi:10.1046/j.1365-294x.1998.00374.x

- Matziris DI (1994) **Genetic variation in the phenology of flowering in black pine.** *Silvae Genet* 43:321-328
- McRae TA, Dutkowski GW, Pilbeam DJ, Powell MB, Tier B (2004) **Genetic evaluation using the TREEPLAN® system.** Paper presented at the IUFRO Joint Conference of Division 2 "Forest Genetics and Tree Breeding in the Age of Genomics: Progress and Future" Charleston, SC, USA, 1-5 November 2004
- Meagher TR, Thompson E (1986) **The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction** *Theor Popul Biol* 29:87-106. doi:10.1016/0040-5809(86)90006-7
- Meuwissen THE, Hayes BJ, Goddard ME (2001) **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 157:1819-1829
- Misztal I, Legarra A, Aguilar I (2009) **Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information.** *J Dairy Sci* 92:4648-4655. doi:10.3168/jds.2009-2064
- Moriguchi Y, Taira H, Tani N, Tsumura Y (2004) **Variation of paternal contribution in a seed orchard of *Cryptomeria japonica* determined using microsatellite markers.** *Can J For Res-Rev Can Rech For* 34:1683-1690. doi:10.1139/x04-029
- Moriguchi Y, Yomogida H, Iwata H, Takahashi M, Taira H, Tsumura Y (2011) **Paternity analysis in a progeny test of *Cryptomeria japonica* revealed adverse effects of pollen contamination from outside seed orchards on morphological traits.** *Tree Genet Genomes* 7:1089-1097. doi:10.1007/s11295-011-0397-z
- Muller-Starck G (1995) **Protection of genetic variability in forest trees.** *Forest Genetics* 2:121-124
- Mullin TJ (2014) **OPSEL 1.0: A computer program for optimal selection in forest tree breeding.** Technical Report Nr 841-2014, Arbetsrapport Från Skogforsk
- Mullin TJ, Hallander J, Rosvall O, Andersson B (2010) **Using simulation to optimise tree breeding programmes in Europe: an introduction to POPSIM.** Skogforsk NR 711.
- Munoz PR, Resende MFR, Huber DA, Quesada T, Resende MDV, Neale DB, Wegrzyn JL, Kirst M, Peter GF (2014) **Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy.** *Crop Sci* 54:1115-1123. doi:10.2135/cropsci2012.12.0673
- Muranty H, Jorge V, Bastien C, Lepoittevin C, Bouffier L, Sanchez L (2014) **Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops.** *Tree Genet Genomes* 10:1491-1510. doi:10.1007/s11295-014-0790-5
- Na SJ, Lee HS, Han SU, Park JM, Kang KS (2015) **Estimation of genetic gain and diversity under various genetic thinning scenarios in a breeding seed orchard of *Quercus acutissima*.** *Scand J Forest Res* 30:377-381. doi:10.1080/02827581.2015.1018936
- Namkoong G, Kang H, Brouard JS (1988) **Tree breeding: principles and strategies. Monographs on theoretical and applied genetics.** 11. Springer-Verlag, New York
- Neale DB, Savolainen O (2004) **Association genetics of complex traits in conifers.** *TRENDS in Plant Science* 9:325-330. doi:10.1016/j.tplants.2004.05.006
- Ng CH, Lee SL, Ng KKS, Lee CT, Muhammad N (2012) **Establishment of full-sib families in *Shorea platyclados* using paternity analysis: an alternative to controlled pollination.** *J For Res* 17:479-487. doi:10.1007/s10310-011-0318-x
- Olsson T, Lindgren D, Li B (2001) **Balancing genetic gain and relatedness in seed orchards.** *Silvae Genet* 50:222-227
- Pâques L (2013) **Forest tree breeding in Europe : current state-of-the-art and perspectives.** ed. / Luc E Pâques. Springer- Managing Forest Ecosystems, Vol. 25. doi:10.1007/978-94-007-6146-9

- Payn T, Carnus JM, Freer-Smith P, Kimberley M, Kollert W, Liu SR, Orazio C, Rodriguez L, Silva LN, Wingfield MJ (2015) **Changes in planted forests and future global implications**. For Ecol Manage 352:57-67. doi:10.1016/j.foreco.2015.06.021
- Pilate G, Pâques M, Leplé J-C, Plomion C (2002) **Les biotechnologies chez les arbres forestiers**. Revue Forestière Française 54:161-180
- Plomion C, Bartholomé J, Lesur I, Boury C, Rodríguez-Quilón I, Lagraulet H, Ehrenmann F, Bouffier L, Gion J, Grivet D, De Miguel M, De María N, Cervera M, Bagnoli F, Isik F, Vendramin G, Gonzalez-Martinez S (2015) **High-density SNP assay development for genetic analysis in maritime pine (*Pinus pinaster*)**. Mol Ecol Resour. doi:10.1111/1755-0998.12464
- Plomion C, Chancerel E, Endelman J, Lamy J-B, Mandrou E, Lesur I, Ehrenmann F, Isik F, Bink M, van Heerwaarden J, Bouffier L (2014) **Genome-wide distribution of genetic diversity and linkage disequilibrium in a mass-selected population of maritime pine**. BMC Genomics 15:171
- Plomion C, Durel CE, Omalley DM (1996a) **Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions**. Theor Appl Genet 93:849-858. doi:10.1007/bf00224085
- Plomion C, Durel CE, Verhaegen D (1996b) **Marker-assisted selection in forest tree breeding programs as illustrated by two examples: maritime pine and eucalyptus**. Ann Sci For 53:819-848. doi:10.1051/forest:19960403
- Pot D, Rodrigues JC, Rozenberg P, Chantre G, Tibbits J, Cahalan C, Pichavant F, Plomion C (2006) **QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.)**. Tree Genet Genomes 2:10-24. doi:10.1007/s11295-005-0026-9
- Ragoussis J (2006) **Genotyping technologies for all**. Drug Discovery Today: Technologies 3:115-122. doi:10.1016/j.ddtec.2006.06.013
- Resende MDV, Resende MFR, Sansaloni CP, Petrolí CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, 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 Phytol 194:116-128. doi:10.1111/j.1469-8137.2011.04038.x
- Resende MFR, 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 190:1503-+. doi:10.1534/genetics.111.137026
- Ribeiro M, Le Provost G, Gerber S, Vendramin G, Anzidei M, Decroocq S, Marpeaud A, Mariette S, Plomion C (2002) **Origin identification of maritime pine stands in France using chloroplast single-sequence repeats**. Ann For Sci 59:53-62. doi:10.1051/forest:2002100
- Ross SD, Bower RC (1989) **Cost-effective promotion of flowering in a Douglas-Fir seed orchard by girdling and pulsed stem injection of gibberelin A4/7**. Silvae Genet 38:189-195
- Sanders K, Bennewitz J, Kalm E (2006) **Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population**. J Dairy Sci 89:315-321
- Santos-del-Blanco L, Climent J, González-Martínez SC, Pannell JR (2012) **Genetic differentiation for size at first reproduction through male vs. female functions in the widespread Mediterranean tree *Pinus pinaster***. Ann Bot 110:1449-1460. doi:10.1093/aob/mcs210
- Schefers JM, Weigel KA (2012) **Genomic selection in dairy cattle: integration of DNA testing into breeding programs**. Animal Frontiers 2:4-9. doi:10.2527/af.2011-0032
- Seido K, Maeda H, Shiraishi S (2000) **Determination of the selfing rate in a Hinoki (*Chamaecyparis obtusa*) seed orchard by using a chloroplast PCR-SSCP marker**. Silvae Genet 49:165-168

- Sénat français (2014-2015) **Faire de la filière forêt-bois un atout pour la France**. Rapport d'information n° 382 (2014-2015) de MM. Alain HOUPERT et Yannick BOTREL, fait au nom de la commission des finances, déposé le 1er avril 2015. <http://www.senat.fr/rap/r14-382/r14-3820.html>.
- Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000) **Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties**. Theor Appl Genet 101:1273-1281. doi:10.1007/s001220051607
- Shimono A, Wang XR, Torimaru T, Lindgren D, Karlsson B (2011) **Spatial variation in local pollen flow and mating success in a *Picea abies* clone archive and their implications for a novel "breeding without breeding" strategy**. Tree Genet Genomes 7:499-509. doi:10.1007/s11295-010-0351-5
- Soliman T, Mourits MCM, van der Werf W, Hengeveld GM, Robinet C, Lansink A (2012) **Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe**. PLoS One 7:12. doi:10.1371/journal.pone.0045505
- Stoehr M, Webber J, Woods J (2004) **Protocol for rating seed orchard seedlots in British Columbia: quantifying genetic gain and diversity**. Forestry 77:297-303. doi:10.1093/forestry/77.4.297
- Stoehr M, Yanchuk A, Xie CY, Sanchez L (2008) **Gain and diversity in advanced generation coastal Douglas-fir selections for seed production populations**. Tree Genet Genomes 4:193-200. doi:10.1007/s11295-007-0100-6
- Telfer EJ, Stovold GT, Li YJ, Silva OB, Grattapaglia DG, Dungey HS (2015) **Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: a comparative analysis of marker data power and robustness**. PLoS One 10:18. doi:10.1371/journal.pone.0130601
- Terray L, Pagé C, Déqué M, Flecher C (2010) **L'évolution du climat en France au travers de quelques indicateurs agroclimatiques**. In Brisson, N., Levrault, F. (Eds) : Livre Vert CLIMATOR, ADEME, 15-30.
- Thivolle-Cazat A, Colin A, Drouineau S, Cavaignac S, Bailly A, Labbe T, Lambert P (2013) **Etude ressource: analyse prospective de la ressource forestière et des disponibilités en bois de la région Aquitaine à l'horizon 2025**. Rapport final. Juillet 2013. [http://inventaire-forestier.ign.fr/spip/IMG/pdf/Rapport\\_final\\_Etude\\_disponibilites\\_Aquitaine.pdf](http://inventaire-forestier.ign.fr/spip/IMG/pdf/Rapport_final_Etude_disponibilites_Aquitaine.pdf).
- Torimaru T, Wennstrom U, Andersson B, Almqvist C, Wang XR (2013) **Reduction of pollen contamination in Scots pine seed orchard crop by tent isolation**. Scand J Forest Res 28:715-723. doi:10.1080/02827581.2013.838298
- VanRaden PM, Tassell CPv, Wiggans GR, Sonstegard TS, Schanabel RD, Taylor JF, Schenkel FS (2009) **Reliability of genomic predictions for North American Holstein bulls**. J Dairy Sci 92:16-24. doi:10.3168/jds.2008-1514
- Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L (2015) **Paternity recovery in two maritime pine polycross mating designs and consequences for breeding**. Tree Genet Genomes 11:13. doi:10.1007/s11295-015-0932-4
- White T (1987) **A conceptual framework for tree improvement programs**. New For 1:325-342. doi:10.1007/BF00031742
- White TL, Adams WT, Neale DB (2007) **Advanced generation breeding strategies - Breeding population size, structure and management**. In: Forest Genetics, CAB Books, Chap 17, p 489. In. doi:10.1079/9781845932855.0479
- Wilcox PL, Carson SD, Richardson TE, Ball RD, Horgan GP, Carter P (2001) **Benefit-cost analysis of DNA marker based-selection in progenies of *Pinus radiata* seed orchard parents**. Can J For Res-Rev Can Rech For 31:2213-2224. doi:10.1139/cjfr-31-12-2213
- Wright S (1922) **Coefficients of inbreeding and relationship**. Amer Nat 56:330-338



- Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale D, McKeand S, Whetten R (2012) **SNP markers trace familial linkages in a cloned population of *Pinus taeda* - prospects for genomic selection.** *Tree Genet Genomes* 8:1307-1318. doi:10.1007/s11295-012-0516-5
- Zapata-Valenzuela J, Whetten RW, Neale D, McKeand S, Isik F (2013) **Genomic estimated breeding values using genomic relationship matrices in a cloned population of Loblolly Pine.** *G3-Genes Genomes Genet* 3:909-916. doi:10.1534/g3.113.005975
- Zas R, Merlo E, Fernandez-Lopez J (2004) **Juvenile-mature genetic correlations in *Pinus pinaster* Ait. under different nutrient x water regimes.** *Silvae Genet* 53:124-129
- Zhao P, Zhang SX, Woeste K (2013) **Genotypic data changes family rank for growth and quality traits in a black walnut (*Juglans nigra* L.) progeny test.** *New For* 44:357-368. doi:10.1007/s11056-012-9343-7
- Zobel BJ, Talbert JT (1984) **Applied Forest Tree Improvement** (p. 528). New York: Wiley.

# Annexe 1: Liste des publications et communications issues de cette thèse

## ➤ Publication scientifique dans une revue avec comité de lecture

Vidal M, Plomion C, Harvengt L, Raffin A, Boury C, Bouffier L (2015) Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. *Tree Genet Genomes* 11:1-13. doi: 10.1007/s11295-015-0932-4

## ➤ Publication scientifique dans une revue avec comité de lecture, en préparation

Vidal M, Plomion C, Harvengt L, Raffin A, Bouffier L (2016) Forward selection in a maritime pine polycross progeny trial using pedigree reconstruction.  
Soumis à *Annals of Forest Science* en janvier 2016.

## ➤ Communications orales

- Vidal M., Plomion C., Harvengt L., Bourry C. and Bouffier L. (2015) Paternity recovery in a maritime pine polycross trial using SNPs: Consequences for breeding (oral communication and poster). In Vettori C. and Vendramin G. G. (eds) Proceedings of the IUFRO Tree Biotechnology conference 2015 (Working Group 2.04.06), Florence, Italy, 7-12 June 2015.

- Vidal M. (2014) « Assessment of polymix breeding », WP5, Meeting annuel du projet européen ProCoGen, Alcalá de Henares, Espagne. Février 2014

- Vidal M. Optimisation des stratégies d'amélioration génétique du pin maritime grâce aux marqueurs moléculaires. Thèses des bois, Pôle de Compétitivité XYLOFUTUR, le 2 juillet 2015. Cestas.

## ➤ Posters

- Vidal M., Ruby M., Alazard P., Harvengt L. and Bouffier L. (2013) Paternity recovery in a maritime pine polycross trial (poster 36). In Proceedings of the Forest Genetics 2013, Integration of Genomics Tools into Breeding Programs, a joint meeting of Canadian Forest Genetics Association, Western Forest Genetics Association, IUFRO Population, Ecological and Conservation Genetics (Working Group 2.04.01) and IUFRO Breeding and Genetic Resources of Pacific Northwest Conifers (Working Group 2.02.05), Whistler, British Columbia, Canada; July 22-25, 2013

- Vidal M., Plomion C., Harvengt L., Bourry C. and Bouffier L. (2015) Paternity recovery in a maritime pine polycross trial using SNPs: Consequences for breeding (oral communication and poster). In Vettori C. and Vendramin G. G. (eds) Proceedings of the IUFRO Tree Biotechnology conference 2015 (Working Group 2.04.06), Florence, Italy, 7-12 June 2015.