



L'étude de caractères complexes en sélection : vers des modèles de réseau de gènes

Léopoldo Sanchez

► To cite this version:

Léopoldo Sanchez. L'étude de caractères complexes en sélection : vers des modèles de réseau de gènes. Sciences du Vivant [q-bio]. Université d'Orléans, 2010. tel-02814402

HAL Id: tel-02814402

<https://hal.inrae.fr/tel-02814402>

Submitted on 6 Jun 2020

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.

UNIVERSITE D'ORLEANS

Rapport présenté par Leopoldo Sánchez-Rodríguez

Chargé de Recherches à l'INRA d'Orléans

Docteur en Sciences Biologiques

En vue de l'obtention de l'Habilitation à Diriger des Recherches

(DIPLOME NATIONAL INSTITUTE PAR L'ARRETE MINISTERIEL DU 23 NOVEMBRE 1988)

L'étude de caractères complexes en sélection :

vers des modèles de réseau de gènes

Composition du jury :

Mme. **Christine DILLMANN**, Professeur, Université Paris-Sud XI, Rapporteur

M. **Gilles CHARMET**, Directeur de recherches, INRA Clermont-Ferrand, Rapporteur

M. **Jean-Marc BOUVET**, Directeur de recherches, CIRAD Montpellier, Rapporteur

M. **Alain FRANC**, Directeur de recherches, INRA Bordeaux, Examineur

M. **François LEFEVRE**, Directeur de recherches, INRA Avignon, Examineur

M. **Martine DECOVILLE**, Maître de conférences, Université d'Orléans, Examineur

M. **Eric LAINE**, Professeur, Université d'Orléans, Examineur

Sommaire

Introduction ... 4

Bilan de mes recherches et leur contexte scientifique ... 5

Perspectives : projets de recherche engagés et futurs ... 27

Liste de projets de recherche ... 31

Collaborateurs ... 32

Références bibliographiques ... 34

Annexe : Curriculum Vitae ... 39

1. Introduction

L'amélioration génétique des espèces forestières est sans doute la solution la plus efficace pour répondre aux besoins quantitatifs et qualitatifs croissants en bois et biomatériaux. Elle est aussi un modèle réduit exceptionnel pour notre compréhension de l'évolution et l'adaptation du vivant. Cependant, et à contre courant dans le monde actuel du *court terme*, elle reste une action de très longue haleine, qui nécessite une continuité d'efforts, et une certaine indépendance face aux effets de mode. Indépendance, certes, mais il faut être attentif toutefois aux nouveaux défis qui se posent, notamment le changement climatique et l'essor des outils en génomique et l'ingénierie du vivant.

Mon projet personnel de recherche s'alimente principalement par ces deux défis : comment les stratégies en amélioration génétique existantes et d'avenir doivent être modelées pour répondre au risque de maladaptation des espèces face aux changements environnementaux et, simultanément, comment intégrer harmonieusement le potentiel offert par les outils en génomique ? Plusieurs outils sont utilisés à ce propos, dont les modèles théoriques et de simulation, et les outils d'analyse quantitative appliqués aux populations expérimentales et naturelles. Les modèles prédictifs en biologie (biologie prédictive) et l'adaptation des populations au changement climatique sont deux parmi les quatre axes de réflexion sur l'orientation scientifique de l'INRA pour la décennie qui s'ouvre devant nous.

Les activités de mon travail font référence aux différentes étapes du processus de gestion et de valorisation des ressources génétiques : l'étude de la structure de la diversité génétique des populations naturelles et recherche des prédicteurs de la valeur adaptative et de la plasticité des génotypes ; l'échantillonnage et la constitution de collections représentatives de la diversité naturelle ; et la réflexion sur l'impact de la sélection artificielle sur la diversité et la variabilité génétiques.

Dans le présent rapport d'HdR, j'ai essayé de refléter une partie de ces différents volets, sans être exhaustif, pour mieux définir la trajectoire de mes recherches. Cette trajectoire pourra sembler écartée d'une linéarité *idéale*, notamment par l'absence d'un modèle d'étude unique. Cela obéit à une certaine richesse d'approches, d'équipes de travail, de projets et même de pays, qui ont parsemé mon cursus. Je présente cette trajectoire dans le bilan suivant, notamment au sein de l'INRA, et dans le contexte scientifique qui correspond à chaque étape. Ce bilan s'ouvrira sur les perspectives, les projets engagés et futurs où la modélisation du lien gène-phénotype via les réseaux de gènes prendra une place prépondérante.

2. Bilan de mes recherches et leur contexte scientifique

L'importance de la diversité génétique en sélection

Qu'est-ce qu'on entend d'abord par une **gestion raisonnée** dans le contexte de l'amélioration génétique ? Une gestion raisonnée est toujours basée sur un compromis entre variabilité génétique et réponse à la sélection, et l'objectif couvre une période de temps étendu ou un nombre de générations important. Dans le cas de populations suivant un régime d'allofécondation, la perte d'une variabilité génétique initiale est naturelle et inévitable ([76]). Cette perte s'explique de la façon suivante : le processus de ségrégation, qui conduit à la formation de gamètes, révèle à chaque cycle une partie de la variation initiale disponible ; et la taille finie de l'échantillonnage gamétique résultante limite la portée en termes de variation génétique pour la génération suivante. L'échantillonnage gamétique entre générations produit la dérive génétique, qui amplifie les différences en représentation de la diversité génétique entre générations. Dans le compromis d'une gestion raisonnée, la perte de diversité génétique doit être maîtrisée au même titre que le gain en réponse à la sélection.

La sélection épuise la variabilité disponible dans la population, en favorisant certaines variantes. Pourquoi donc est-il si important de maîtriser une perte qui de toute façon est inévitable ? Il faut d'abord noter d'une part que la sélection n'est pas un processus précis, infaillible, loin de là. L'effet d'un gène, même favorable, peut passer inaperçu parmi les effets d'autres gènes, parmi les interactions entre gènes, et entre eux et l'environnement. D'autre part, la taille finie des populations hétérozygotes limite pratiquement à néant le nombre de gamètes concentrant la totalité d'allèles favorables pour un caractère quantitatif. Le résultat est que, bien que la perte de variation reste inévitable dans une population à taille finie, le taux auquel cette perte se produit est variable et dépendant d'une interaction complexe entre la recombinaison et la sélection. Il y a certaines règles simples, déduites de la théorie de la taille efficace, qui permettent de limiter cette perte ([13]). Ainsi, l'élimination de fluctuations éventuelles, entre générations dans la taille de la population de reproducteurs et dans la taille de familles résultant, peut réduire le taux de perte dû à la dérive génétique. Même si elles sont efficaces, ces mesures restent difficiles à mettre en œuvre dans la pratique, du fait de l'antagonisme avec d'autres contraintes de type biologique ou de gestion.

Avec une approche très simplificatrice mais révélatrice, on peut conclure que la dérive génétique aurait trois conséquences défavorables dans les populations améliorées. D'abord, une perte de diversité génétique favorable due à la faillibilité de la sélection et à la taille finie de la population sélectionnée, et par conséquent une réduction du gain génétique à plus long terme. Une deuxième conséquence serait la perte de la diversité génétique qui n'est pas liée fonctionnellement aux caractères directement ciblés par la sélection artificielle. Cette diversité pourrait, toutefois, se révéler nécessaire pour l'adaptation de la population à des circonstances autres que celles propres à l'amélioration : changement climatique, nouvelles maladies, etc. Cette perte est déterminée par la taille finie de la population sélectionnée. Une troisième conséquence, pas moins importante, serait la fixation des variantes génétiques indésirables. Indésirables, parce que leur effet réduit la valeur pour les caractères ciblés par la sélection artificielle, ou parce que la réduction prend effet sur la valeur adaptative des individus sélectionnés. Souvent on dénomme ces variantes comme des allèles délétères ou des létaux. Dans les deux cas, on parle généralement de dépression de

consanguinité, le phénomène par lequel la valeur phénotypique d'une population résulte réduite avec l'augmentation de l'homozygotie. Ces risques qu'on vient de citer justifient généralement le contrôle explicite de la dérive génétique et de la perte de diversité dans les programmes de sélection et de conservation.

Ce contrôle explicite de la perte de variation est dans la plupart des programmes basé sur une mesure *consensuelle* de taille efficace, ou de taux maximal d'augmentation de la consanguinité. Notre connaissance sur les niveaux de diversité favorable et nuisible dans les populations domestiquées est très limitée et indirecte. Certaines expériences de consanguinité forcée montrent les effets de la dépression sur des caractères productifs ou d'autres caractères liés à la valeur de la fitness (reproduction, survie, etc). L'utilité de ces expériences réside dans la définition de limites à ne pas dépasser pour la consanguinité, avec des taux consensuels qui résultent de placer cet horizon pour la consanguinité et de diviser la période résultante dans un nombre de cycles de sélection *suffisamment* raisonnable. Pour illustrer cette situation nous prenons l'exemple d'une espèce domestique devenue référence en amélioration : la race de bovin laitier Holstein. Ainsi, une taille efficace de 80 a été considérée raisonnable pour la population mondiale de cette race ([23]), suffisant pour garantir un gain génétique à long terme avec des taux de consanguinité qui ne devraient pas dépasser les 0.5% par an. D'autres auteurs suggèrent des tailles comparables, entre 30 et 120, sans cibler une population ou une espèce particulière ([1]). Malgré cela, le taux réel de consanguinité chez la race Holstein a fortement augmenté à partir des années 1990, avec la moitié des nouveaux mâles sélectionnés pour la reproduction provenant de 10 meilleurs taureaux, atteignant des valeurs de 1.3% par génération (voir [74], [15]). Cette même situation est applicable à d'autres espèces d'animaux domestiques de large diffusion et ayant des programmes d'amélioration à grande échelle.

Les espèces de l'amélioration forestière : un cas particulier ?

Pouvons-nous nous inspirer de grands programmes d'amélioration animale pour la gestion raisonnée des programmes chez les espèces forestières ? Certainement oui. Chez la plupart des espèces forestières améliorées, la situation peut sembler beaucoup plus bénigne que celle décrite pour le bovin. D'abord, même les populations plus avancées en amélioration sont encore proches des populations non domestiques, dont les tailles efficaces sont souvent considérées énormes ($\geq 10^4$) pour la plupart des espèces d'intérêt ([10], [45]). La biologie de ces espèces, pensons aux conifères de l'hémisphère nord, avec des aires de distribution vastes et continues, avec une forte tendance à l'allofécondation, et aux flux gamétiques importants entre populations voisines, peut expliquer aisément l'ampleur de cette diversité ([29]). Cependant, cette riche diversité pourrait être apparente, au moins en termes relatifs. Ainsi, la diversité moléculaire pour la plupart des espèces de conifères d'intérêt industriel montre des niveaux comparables ou même inférieurs à ceux d'autres espèces de plantes moins ubiquistes ([57]). Une évolution neutre *parcimonieuse*, caractéristique des arbres, est parmi les raisons suggérées. Elle conduirait à des situations de déséquilibre *chronique* face à des perturbations environnementales plus rapides.

Mais si ces hypothèses autour de la diversité de nos forêts sont encore débattables (voir [45]), d'autres caractéristiques propres aux espèces forestières améliorées sont par contre incontestables, avec des effets potentiels majeurs sur la préservation de leur diversité. Une de ces caractéristiques est la capacité reproductive, laquelle peut permettre des taux de sélection considérables. L'histoire

du NZ55 illustre l'extrême du possible ici. Derrière NZ55 se cache un exemplaire exceptionnel de pin de Monterey (*Pinus radiata* D. Don) très apprécié par les améliorateurs de l'espèce. Cet individu est représenté dans la plupart des vergers à graines de *Pinus radiata* de la Nouvelle Zélande et de l'Australie depuis 1957. Dag Lindgren (Université Suédoise de Sciences Forestières) a estimé à 50 millions le nombre de descendants vivants qui ont été obtenus de ses graines (données non publiées). Cette cohorte représente un volume de bois équivalent à la production annuelle de la Finlande, toutes espèces confondues. Mais cet exemple n'est pas isolé dans son exceptionnalité, et d'autres espèces ont aussi leur « NZ55 ». Une variété de l'espèce *Populus nigra*, le génotype *italica*, est littéralement omniprésente dans le paysage de la plupart des continents, avec une probabilité non négligeable d'introgresser ses gènes dans les peuplements naturels. Ces exemples montrent simplement que la capacité reproductive d'un seul arbre peut être multipliée énormément de façon artificielle, largement au-delà des capacités maximales observées dans les programmes d'insémination artificielle des bovins (jusqu'à 3000 descendants mâles et 250000 femelles pour un taureau de l'élite).

Le risque d'une contribution génétique déséquilibrée peut être combiné à une autre particularité présente dans la plupart des espèces forestières améliorées, et notamment chez les conifères : un niveau considérable de ce qu'on appelle **fardeau génétique**. Le fardeau génétique désigne l'ensemble des mutations génétiques défavorables dans une population, celles responsables in fine de la dépression de consanguinité. Suite à des expériences de consanguinité forcée, comprenant notamment des autofécondations, le niveau du fardeau génétique peut être inféré indirectement à partir du nombre d'équivalents létaux. Chaque équivalent léthal aurait l'effet en état d'homozygotie de provoquer la mort à son porteur, et cette mesure s'interprète comme un nombre efficace de gènes. Les conifères montrent des valeurs d'équivalents létaux normalement supérieures à 5 ([47], [75]). Les régimes reproductifs en allofécondation des conifères et les grandes tailles de population facilitent le maintien d'un tel niveau. D'autres taxons comprenant animaux et plantes annuelles, montrent typiquement des valeurs inférieures à 5. Tout cela illustre que, bien que les programmes d'amélioration forestière aient accès encore à une diversité génétique considérable, certainement plus importante que celle des races d'animaux domestiques, une gestion raisonnée de ces ressources devrait être probablement beaucoup plus conservatrice dans ces objectifs que les schémas d'amélioration pour le bovin, où les taux de sélection potentiels et le fardeau génétique restent moins importants que dans la plupart des espèces forestières améliorées. Ce sont les méthodes de gestion raisonnée, développées initialement pour contenir la dérive et la consanguinité en génétique animale, qui ont un intérêt dans la génétique forestière.

Les stratégies de sélection

Il y a une pléthore de stratégies qui permettent un contrôle plus ou moins explicite de la perte de diversité due à la dérive génétique ou à la consanguinité. Sans vouloir trop simplifier, on peut classer ces stratégies en deux types : celles qui agissent sur le schéma de sélection et de croisements (nommée ci-après stratégies **de planification**), et celles qui agissent sur le critère de sélection (ci-après **d'évaluation**). Les premières sont plus simples conceptuellement, mais ne permettent pas en général un contrôle explicite sur le niveau de diversité résultante. Celui-ci doit être inféré indépendamment de façon théorique. Comme exemples de stratégies de planification, on peut citer l'égalisation des contributions des parents en nombre de descendants ([73]), ou encore, les croisements factoriels ([64]). La plupart de programmes avancés de sélection ou de conservation

comportent dans une certaine mesure des stratégies de planification. Mon apport à l'étude de ce type de stratégies se résume dans un travail ([53]) qui compare par simulation différents degrés de déséquilibre entre la contribution de parents à la génération de descendants. L'originalité de ce travail réside dans l'utilisation d'une approche de coalescence permettant d'inférer la perte d'allèles fondateurs, et ainsi visualiser l'effet in fine d'une réduction éventuelle en taille efficace.

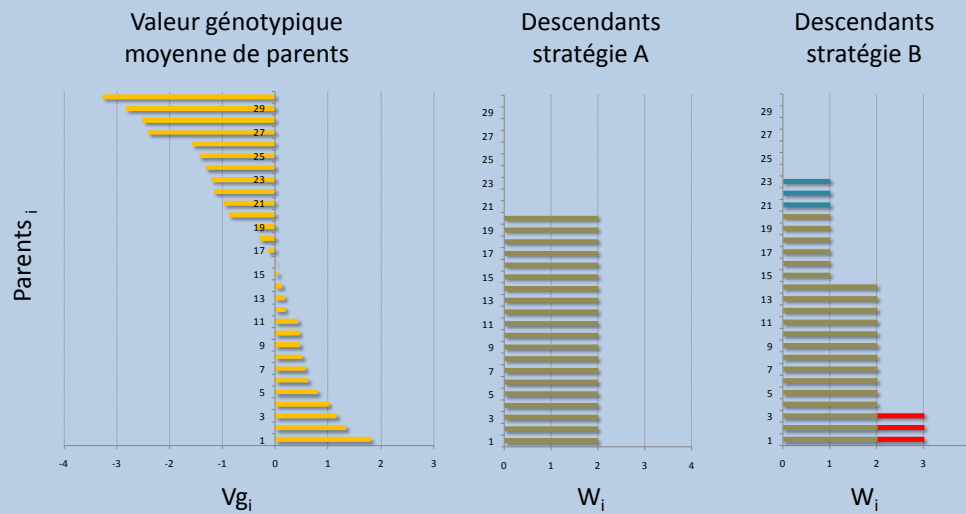
Parmi les stratégies d'évaluation, donc agissant sur le critère de sélection, on trouve à nouveau deux groupes qui se différencient dans la manière de prendre en compte la perte de diversité : soit implicitement, soit explicitement et indépendamment de la performance à évaluer. Ainsi, on peut pondérer l'information familiale et individuelle pendant la phase d'évaluation des candidats de telle façon que le résultat de la sélection conduit à une réduction implicite de la dérive génétique. Bien que les apports de l'information familiale augmentent l'exactitude de l'estimation de la valeur reproductive des candidats (donnons ici l'exemple de l'évaluation BLUP, « *Best Linear Unbiased Predictor* » ou meilleur prédicteur linéaire sans biais), ils produisent aussi une augmentation de la probabilité de sélectionner des individus génétiquement apparentés. Une façon simple de contrebalancer le poids de l'information familiale consiste à considérer pour l'évaluation des candidats à la sélection une héritabilité artificiellement supérieure à la meilleure estimation disponible ([46]).

Cependant, et non sans surprises, les stratégies d'évaluation les plus performantes en gain génétique par unité de diversité sont celles basées sur l'optimisation numérique, et guidées par une contrainte explicite sur la perte de diversité. Ces méthodes, disons optimales, ont une racine commune qui se trouve dans le concept de sélection pondérée (« *weighted selection* », [69]). Dans la **boîte 1** ci-dessous, on illustre le principe de cette stratégie avec un exemple numérique simulé.

Selon cette stratégie, la taille efficace (N_e) et le gain génétique à un moment donné, dans une certaine mesure antagonistes, sont fonction d'une même variable de décision : la contribution familiale, laquelle peut donc faire l'objet d'une optimisation. Cette méthode a été fondamentale pour le développement d'autres méthodes de sélection optimisée plus sophistiquées. Mais, avant de franchir le pas vers ces développements nouveaux, on peut commencer avec un exemple d'implémentation de la sélection pondérée dans une espèce forestière ([2]). Le cas concerne une des populations du programme d'amélioration multi-populations de Pin sylvestre en Suède, où la question était de savoir comment bâtir une deuxième génération de sélections performante, tout en restant dans le critère de diversité propre aux programmes suédois. Ces critères de diversité prennent comme référence le résultat d'une sélection exclusivement intrafamiliale, avec des contributions par famille approximativement équitables, ce qui correspond à notre stratégie A dans l'exemple de la boîte 1. L'alternative de la sélection pondérée était donc potentiellement intéressante avec un choix approprié de contraintes. La nouveauté, par rapport à l'alternative classique de sélection pondérée (**boîte 1**), était l'utilisation d'une taille efficace décrite en fonction de la moyenne de la matrice d'apparement entre les individus sélectionnés. Celle-ci prend en compte de façon explicite l'effet de la variation des contributions et de la dérive génétique depuis le moment de la fondation du pedigree. D'autres auteurs avaient déjà proposé cette alternative pour la taille efficace et montré son efficacité par simulation ([14]). Nos résultats avec la population de pin montrent un avantage significatif en gain génétique par rapport à la stratégie de référence, entre 3 et 8% en fonction des contraintes et du régime de croisements choisis. Cette alternative est considérée maintenant sérieusement par les gestionnaires du programme suédois, et a incité de nouveaux

travaux locaux sur le sujet, notamment une thèse récente ([26]). Le même principe a été appliqué à

1. Sélection pondérée (« *weighted selection* »)



Gain génétique

$$\sum [Vg_i \times W_i] / N$$

0.41

0.47

Taille efficace

$$[\sum (W_i^2 + W_i) / (2 \times N^2)]^{-1}$$

26.67

26.67

La **sélection pondérée** en action pour une population de 30 paires de parents. Le graphique en haut à gauche représente la distribution de valeurs génotypiques moyennes par paire, une distribution approximativement normale. Typiquement, la stratégie de référence comprend un nombre égal de descendants par paire de reproducteurs (nommée **stratégie A**). La stratégie alternative (ici **B**) donne un nombre inégal de descendants par paire en fonction de la valeur génotypique de la paire, et en prenant plus de paires que A mais sans modifier le taux de sélection (toujours 40 descendants). Le gain génétique est la moyenne de valeurs génotypiques, et la taille efficace est une simplification basée sur la moyenne de l'apparentement. Utilisant une technique de programmation quadratique (optimisation), on peut trouver le vecteur W qui maximise le gain génétique en observant une restriction sur N_e (en l'occurrence, égal à celle de A).

Dans l'exemple montré ici, B obtient un gain génétique 15.8% plus important que A, sans modifier la taille efficace. Dans le graphique de droite, avec la distribution de descendants sous B, les trois paires de parents les plus performants reçoivent une contribution plus importante (en rouge), laquelle est responsable du gain extra. Pour compenser cet excès d'apparentement, d'autres familles supplémentaires sont mises à contribution (en bleu foncé). Bien entendu, le même principe pourrait être appliqué pour maximiser N_e en observant une contrainte sur le niveau de gain génétique.

la population d'amélioration du Douglas au Canada, où nous avons démontré la supériorité par rapport aux systèmes sans contraintes explicites ([65]).

Bien que ce soit la sélection qui ait typiquement l'impact le plus fort sur la diversité de la population, le régime de croisements ne doit pas être négligé comme facteur d'importance, pouvant aussi

affecter profondément la dynamique de la diversité. En fait, parmi les nouveaux perfectionnements sur la base de la sélection pondérée, on trouve la prise en compte du régime de croisements. Souvent, dans les populations d'amélioration, les individus élites sont issus de croisements contrôlés, dont les deux parents participants sont connus. Le choix de parents par l'améliorateur suit souvent des critères de performance, par exemple, les individus croisés partagent des niveaux phénotypiques pour le caractère d'intérêt comparables. Ceci, nommé homogamie, constitue d'ailleurs le processus à l'origine de la plupart des races d'animaux domestiques ([18]). Dans mon travail cité auparavant ([2]), les régimes de croisements choisis appartenaient à cette dernière catégorie de croisements homogamiques. Cependant, le croisement entre individus suivant une ressemblance phénotypique a généralement comme conséquence une augmentation de la consanguinité, sur la base qu'au moins une partie de la ressemblance phénotypique a des causalités génétiques. Sur le même principe, un plan de croisements qui évite les couples d'individus apparentés réduit systématiquement la consanguinité des descendants et préserve la diversité dans la population. Cet outil est souvent préconisé pour la conservation de la diversité génétique dans le cadre des programmes d'amélioration et de conservation à petite taille efficace, sous la forme de minimisation de l'apparentement ([14]).

Une suite logique semblerait l'intégration de la *sélection pondérée*, laquelle contrôle la contribution en descendants de parents, et un régime de croisements qui minimise l'apparentement entre parents. De cette façon, les deux étapes clefs du processus d'amélioration, sélection et croisements, seraient coordonnées simultanément pour une réduction potentiellement plus efficace de la dérive génétique, et par rapport à chaque étape considérée séparément. En fait, cette combinaison prometteuse a été déjà proposée sous le nom de sélection de croisements (« *mate selection* ») ([67]). Avec ce système, la sélection évalue une série de croisements parmi tous les croisements possibles entre les individus candidats. Ainsi, la différence entre deux séries de croisements quelconques évaluées par le système peut résider dans le nombre de parents impliqués, dans la distribution de contributions en descendants de ces parents ou dans les croisements. Le choix de la série optimale doit maximiser le gain génétique dans la génération des descendants tout en observant une contrainte sur la perte de diversité. Plusieurs travaux théoriques (par exemple, [9], [68], [77]) suggèrent que ce système peut apporter des bénéfices substantiels, avec plus de gain génétique, plus de taille efficace ou les deux bénéfices simultanément, suivant la définition du problème et de la contrainte. Une quantification de ce bénéfice théorique serait entre 3 et 10% en réponse à la sélection, ou entre 10 et 20% en taille efficace, considérant une variété d'assumptions génétiques, de tailles et de taux de sélection parmi les études considérées. Cet ordre de grandeur non négligeable a motivé l'engagement sur des approches expérimentales, beaucoup plus lourdes mais réalistes, en considérant surtout des organismes modèles comme la souris ou la Drosophile. J'ai présenté un article de révision ([51]) qui résume ces expérimentations, et il conclut que les bénéfices réels sont en général en concordance avec les prédictions théoriques, mais une variation importante a été constatée entre les résultats des études, peut-être fruit du faible nombre de répliques et d'une certaine hétérogénéité expérimentale qui est absente dans les études théoriques.

Depuis la proposition de la *sélection pondérée* et la *sélection de croisements*, plusieurs nouveautés ont été introduites concernant l'optimisation des stratégies d'amélioration, notamment une formulation plus générale de la sélection optimale, basée sur le concept de contribution génétique à long terme, et un fort développement de la théorie prédictive pour le gain génétique et la taille efficace en sélection. Dans les paragraphes suivants, je présente brièvement certains aspects du

nouveau cadre théorique pour prédiction de la variabilité en sélection, et comment ces avancements théoriques ont finalement permis une généralisation du concept de sélection optimale.

Prédire pour mieux gérer

A travers les exemples de stratégies de sélection précédents, nous avons eu un aperçu rapide de procédures prévues pour des situations réelles : typiquement la prise de décisions concernant la sélection à une génération donnée, ou la gestion des croisements une fois que le pollen est disponible pour l'améliorateur. Cependant, certaines décisions doivent être prises en amont, avant la praxis, là où il faut bâtir des recommandations générales avant même la constitution des programmes de conservation ou d'amélioration. Ces décisions ont reposé jusqu'à très peu sur un cadre théorique robuste, dans certaines limites, et assez simple. Les prédictions de gain génétique résultent ainsi de l'estimation d'un changement linéaire en moyenne phénotypique entre parents et descendants, et en supposant que les autres descripteurs populationnels, notamment variation et forme de la distribution de la population sélectionnée, restent stables. Ces assumptions dérivent d'un **modèle infinitésimal**, duquel on reparlera plus tard. Toutefois, la théorie et l'expérience montrent que le processus de sélection a la capacité de changer la variation entre générations. D'abord, la base génétique sous-jacente à cette variation comprend raisonnablement un nombre fini de gènes qui suivent les pressions sélectives et la dérive en modifiant sensiblement leurs fréquences et, in fine, la variation génétique résultante. Ensuite, les populations de taille finie ne représentent qu'une partie des combinaisons génotypiques possibles selon leur pool génétique, et ce manque de représentativité se traduit par une variation génétique différente de celle attendue uniquement à partir des fréquences des gènes : le résultat de ce qu'on appelle le déséquilibre de liaison. Certaines approximations corrigent convenablement ces biais pour le court terme, notamment pour la dérive et le déséquilibre de liaison ([42]), mais nous ne traiterons pas ici de ces dérivations.

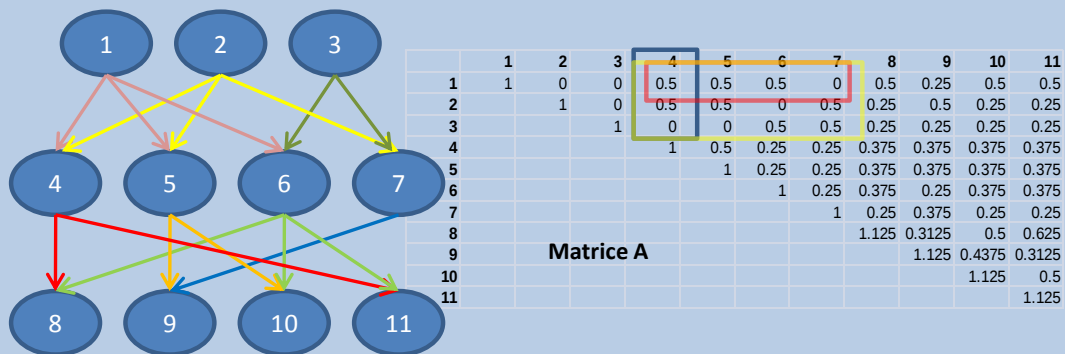
D'autres facteurs encore peuvent affecter considérablement ces prédictions de gain génétique, même à très court terme, comme la consanguinité, la dérive, les effets environnementaux corrélés avec le caractère d'intérêt ou d'autres caractères génétiquement corrélés. Parmi ces facteurs nous mettons en évidence ici le caractère non aléatoire de certains régimes de croisements, les croisements consanguins ou les croisements minimisant l'apparentement. On a aperçu précédemment, l'importance que les régimes non aléatoires peuvent avoir dans la gestion de la diversité à long terme. Toutefois, la littérature scientifique est étonnamment pauvre au sujet des prédictions de gain génétique et de taille efficace en fonction du régime de croisements. Des dérivations satisfaisantes existent pour la taille efficace ([56]), où la variation de fréquences alléliques est modélisée pour un régime particulier de croisements : comme fonction de la proportion de couples frère-sœur par rapport aux couples totaux. Nous avons développé un modèle prédictif plus général, en fonction de l'écart par rapport à un régime de croisements aléatoires, qui explique la dynamique de la variation génétique pour le caractère soumis à sélection, la base du gain génétique, et la taille efficace ([54]). Cet écart dans le régime de croisements est mesuré par une corrélation entre effets génétiques (généralement nommé avec la lettre grecque α) : la corrélation entre les allèles homologues présents dans les deux gamètes qui constituent un individu, conceptuellement équivalente à l'index de fixation F_{IS} de Wright en génétique de populations ([78]). Ainsi, un α positif indique des croisements entre individus qui partagent plus de gènes, éventuellement identiques par descendance, que l'attendu entre individus échantillonnés aléatoirement. Selon un modèle prédictif infinitésimal, la valeur génétique d'un individu donné est composée de deux parties assumées

2. Contribution génétique à long terme

Définition : la contribution génétique à long terme (r) d'un individu i appartenant à la génération t_0 est la proportion de gènes portés par les individus de la génération t_1 et dérivés de i par descendance successive, ou $t_1 - t_0 \rightarrow \infty$. Pour illustrer ici la dérivation de la contribution génétique à long terme, nous prendrons quelques générations seulement, le principe restant le même, et nous établirons les liens entre r et les autres descripteurs classiques de la dynamique d'un pedigree.

Considérons ici premièrement un parent i et sa cohorte de descendants directs (fils et filles), la moitié des gènes présents dans chacun de ses descendants provient du parent i , et en assumant que les autres parents de cette cohorte, les partenaires reproductifs de i , ne sont eux-mêmes pas descendants de i . La contribution génétique de i à cette génération de descendants directs sera de 0.5. Considérons maintenant une deuxième génération non chevauchante, où les individus de la cohorte précédente ont été croisés équitablement avec les individus d'une deuxième cohorte, en assumant aussi que les parents de deux cohortes sont tous des individus non apparentés. La contribution génétique de i à cette génération de ses petit-fils sera de 0.25.

Ces niveaux de 0.5 et 0.25 sont aussi les valeurs correspondant au double de la **probabilité d'identité par descendance** entre parent et enfant et grand parent et petit-fils, respectivement. Ces éléments sont contenus dans la matrice d'apparement additive (**A**) qui donne la covariation génétique additive entre individus d'un même pedigree et partageant des allèles identiques par descendance. Prenons le pedigree suivant, et calculons tous les éléments de la matrice **A**:



Chaque élément d'une colonne de **A** donne la contribution génétique de différents ancêtres fondateurs au génome de l'individu de la colonne. Ainsi les parents et fondateurs 1 et 2 ont chacun une contribution génétique de 0.5 au génome de leur fils 4 (boîte bleue). Prenons maintenant une file correspondant à un fondateur, la somme des éléments appartenant à une même génération (boîte rouge) divisée par la somme de tous les éléments pour cette même génération pour tous les fondateurs (boîte jaune) donne la contribution génétique de ce fondateur, en l'occurrence 3/8 pour 1 (3/8 pour 2 et 1/4 pour 3). Le même exercice pour la génération suivante donnera 0.44 pour 1, 0.31 pour 2 et 0.25 pour 3. Il faut noter que la somme des éléments de la boîte jaune (ou l'équivalent dans les générations successives) est toujours égale au nombre d'individus dans la génération correspondante. La somme des contributions génétiques des fondateurs pour une génération donnée est toujours de 1. Au fil des générations, les contributions génétiques se stabilisent et reflètent la fitness des fondateurs.

indépendantes : la moyenne des valeurs génétiques additives de ses deux parents (moyenne parentale) et un écart dû à la ségrégation ou écart Mendélien. L'originalité du model alternatif proposé ([54]) réside dans la quantification de la *dépendance* entre ceux deux termes, laquelle est due aux effets conjoints de la sélection et du régime de croisements caractérisé par α . En effet, la sélection induit une covariation négative entre la moyenne parentale et l'écart Mendélien, d'une façon équivalente au déséquilibre de liaison entre loci. Toutefois, une covariation négative

équivalente est aussi générée entre la moyenne parentale d'un individu et l'écart Mendélien de son partenaire reproductif, et modulée par la valeur de α . In fine, la variation génétique après sélection résulte inférieure via ces corrélations négatives aux valeurs prédites précédemment avec les modèles classiques, même dans le cadre d'un régime de croisements aléatoires. Cette étude apporte finalement une validation additionnelle de la théorie des contributions génétiques à long terme, outil pour prédire les évolutions de la consanguinité et de la variabilité génétique, en la généralisant à toute modalité d'accouplement entre reproducteurs. Cette théorie des contributions génétiques mérite quelques mots ici, pour mieux comprendre ensuite les dernières applications pour l'optimisation des schémas de sélection et d'accouplements. La définition et la dérivation de la contribution génétique à long terme d'un individu fondateur sont données dans la **boîte 2** de la page précédente.

La théorie des contributions génétiques a été largement développée et généralisée pour les générations chevauchantes et les régimes de sélection basés sur l'utilisation du BLUP ([7], [6], [8]). Ce cadre théorique a changé substantiellement la façon de voir l'accumulation du gain génétique et la perte de diversité, ces deux aspects étant classiquement dérivés de façon indépendante. Brièvement, le gain génétique est selon cette théorie une fonction des succès sélectifs, lesquels sont dépendants de la supériorité génétique des candidats. Les contributions génétiques décrivent explicitement ce succès sélectif sous la forme des lignages de gènes partagés au travers des générations. Parallèlement, ce partage de gènes entre cohortes et entre générations est à la base de la dérivation conceptuelle de l'accumulation de la consanguinité au fil des générations. La théorie des contributions génétiques a bâti le pont conceptuel entre la génération du gain génétique et la perte d'hétérozygotie. Dans la pratique, ce lien a permis notamment de résoudre de façon déterministique l'optimisation a priori du gain génétique sous la contrainte de l'évolution de la diversité génétique, et de proposer des procédures de sélection dynamiques pour maximiser le gain génétique tout en limitant l'augmentation du taux de consanguinité à une valeur fixée a priori ([41]).

S'il fallait citer une application dans laquelle la théorie des contributions génétiques a été particulièrement révélatrice par sa puissance explicative, cela serait sans doute la minimisation de la consanguinité dans les schémas de sélection et de conservation. Ce sujet a nourri une littérature considérable. Nous ne ferons pas justice en citant quelques travaux ici, mais ils ont été significatifs dans le contexte de la théorie de la taille efficace (par exemple, [13], [56]), et dans le développement d'outils de gestion (voir notamment [62], [63]). Parmi ces outils, l'optimisation appliquée à la réduction des taux de consanguinité par génération occupe une place prépondérante, et nous avons déjà introduit le concept de sélection pondérée pour laquelle la minimisation de la consanguinité n'est qu'une variante où le poids donné au gain génétique est négligeable. En étant efficace, le but opérationnel de l'optimisation néglige le mécanisme sous-jacent à la minimisation de la consanguinité. Toutefois, la limite inférieure pour le taux de consanguinité n'avait pas été démontrée mathématiquement dans un cadre général, et l'efficacité des outils d'optimisation restait encore à prouver. Partant uniquement de concepts de la théorie des contributions génétiques, nous avons identifié un schéma de développement d'un pedigree qui atteint la valeur minimale de consanguinité par génération ([48]), et nous avons dérivé mathématiquement ce taux de consanguinité minimale comme fonction uniquement du nombre de reproducteurs. Cela permet de pondérer véritablement l'efficacité des outils d'optimisation pour la minimisation de la consanguinité. Dans la **boîte 3**, j'explique le principe de ce schéma.

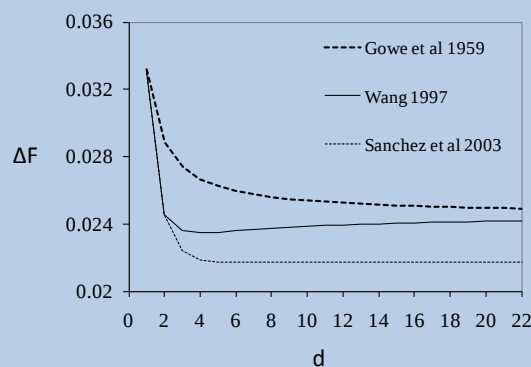
3. Comment atteindre la valeur minimal de ΔF ?

Définition du taux de consanguinité: le taux de consanguinité (ΔF) peut être estimé en fonction des contributions génétiques (r) : $\Delta F = \frac{1}{4} \sum r_i^2$, ou $i = \{1, \dots, N\}$ pour une population de N individus.

Objectif: Pour minimiser la consanguinité à chaque génération, la solution consiste à trouver la valeur minimale de la somme de carrés $\sum r_i^2$, en observant simultanément la contrainte $\sum r_i = 1$. En considérant que la valeur attendue du carré des contributions, $E(r^2)$, est égale à $E(r)^2 + \sigma_r^2$, et que $E(r) = 1/N$, le problème se réduit à minimiser la variance des contributions génétiques, σ_r^2 .

La solution classique la plus simple est quand la population comporte un nombre égal de femelles reproductives (F) et de mâles reproducteurs (M), $M = F = \frac{1}{2}N$, et que chaque paire contribue à la génération suivante exactement avec deux descendants. Ainsi, la contribution génétique de chaque reproducteur est $\frac{1}{N}$, $\sigma_r^2 = 0$ et $\Delta F = \frac{1}{4N}$. Ce taux est la valeur la plus basse atteignable, mais seulement si $M = F$. Des différences entre le nombre de reproducteurs de chaque sexe, typiquement $M = d \times F$ et $d > 1$, impliquent automatiquement que : 1) M femelles contribueront avec un mâle, et 2) $F - M$ femelles contribueront avec une femelle. Par conséquent, la contribution génétique des mères de mâles sera supérieure à la contribution génétique des mères de femelles, et logiquement $\sigma_r^2 > 0$ et $\Delta F > \frac{1}{4N}$. Toute incertitude sur la contribution génétique des descendants contribue à magnifier ΔF .

Nous avons proposé un schéma où les contributions génétiques sont gérées sans incertitude au travers des générations, et ayant pour résultat la valeur de ΔF la plus basse atteignable quand $M \neq F$. Dans ce schéma, il y a $d + 1$ catégories reproductives: **[1]** M parents mâles produisant 1 mâle et d femelles descendants; **[2]** M parents femelles produisant 1 descendant mâle; **[3]** M parents femelles produisant 1 descendant femelle de la catégorie 2; ... ; **[d]** M parents femelles produisant 1 descendant femelle de la catégorie $d - 1$; **[d+1]** M parents femelles produisant 1 descendant femelle de la catégorie d et 1 descendant femelle de la catégorie $d + 1$. Chaque mâle est croisé avec une femelle de chaque catégorie, et les rôles reproductifs de chaque descendant sont prescrits automatiquement *ad infinitum* le long des lignages et sans incertitude. Comme résultat, les descendants avec une contribution génétique plus importante, en l'occurrence les mâles, sont distribués de la façon la plus égalitaire possible entre les femelles reproductives et au fil des générations. Le ΔF minimal résultant est égal à $\phi/16M$, ou $\phi = \frac{4}{3}[1+2(\frac{1}{4})^d]$. Dans le graphique suivant, nous comparons le ΔF de ce système avec d'autres systèmes de minimisation classiques, mais comportant des incertitudes dans l'assignement des contributions génétiques à chaque génération.



Sans trop de risques, on peut affirmer que le corpus de prédictions pour le gain génétique et la consanguinité d'un côté, et les outils d'aide à la sélection en temps réel de l'autre, ont atteint leur maturité dans le cadre d'une sélection basée sur le modèle de caractères polygéniques, cet-à-dire, les caractères supposés être le résultat de nombreux gènes sous-jacents, et pour lesquels la seule source d'information est le phénotype intégrateur. On reviendra plus tard vers ce paradigme, mais maintenant nous allons rentrer dans les enjeux qui résultent de l'identification de gènes particuliers,

comme source additionnelle ou prépondérante d'information sur les caractères quantitatifs soumis à sélection.

Quand la sélection s'approche des gènes

Les développements récents en génomique ont permis le génotypage à grande échelle et la localisation des polymorphismes associés à des caractères productifs et adaptatifs. Cette révolution ouvre certainement de nouvelles perspectives pour l'amélioration génétique forestière, où l'évaluation génétique des individus représente un facteur limitant considérable dans les schémas classiques. Ainsi, dans la plupart des espèces forestières d'intérêt en amélioration génétique, l'évaluation des candidats à la sélection repose substantiellement sur des dispositifs de descendance, souvent issus des familles de demi-frères. Toutefois, une maturité sexuelle généralement tardive, propre à la biologie des espèces longévives comme les arbres forestiers, rallonge inévitablement l'intervalle entre évaluation génétique et croisements. Additionnellement, la grande taille générale des arbres et l'extension des dispositifs d'évaluation limitent le nombre des mesures détaillées et leur rythme d'acquisition par l'améliorateur. Ces particularités placent les espèces forestières comme des candidats idéaux pour une sélection assistée par marqueurs (SAM) ou par gènes (SAG), en vue de réduire la dépendance dans le phénotype. Les bénéfices potentiels en gain génétique (ΔG) peuvent être illustrés avec l'équation suivante, déjà connue comme celle de l'améliorateur :

$$\Delta G = i R \sigma_g / T ;$$

où i est l'intensité de sélection, R la précision de la sélection ou la corrélation entre l'index d'évaluation et la valeur reproductive des individus, σ_g l'écart type génétique, et T l'intervalle de génération qui sépare deux événements de sélection successives. Pour une intensité de sélection donnée, normalement fixée par les contraintes de gestion d'un programme d'amélioration, et un niveau aussi limité de variabilité génétique, c'est le gain en précision de sélection et en temps qui peut apporter un avantage additionnel en ΔG par cycle de sélection. Dans ce sens, la possibilité de cibler directement des mutations causales connues ou marquées à effets significatifs permet d'accroître la précision de la sélection et de raccourcir l'intervalle de génération. Cependant, cette même augmentation de la précision de sélection alimente aussi des questionnements qui concernent l'impact éventuel sur la diversité des populations améliorées résultantes. En théorie, la SAM (et la SAG) permet un accroissement des gains génétiques à court terme, via la fixation rapide des mutations causales ciblées directement par le marquage. En revanche, un poids excessif sur ce gain à court terme pourrait vraisemblablement avoir des conséquences adverses sur la diversité non ciblée par le marquage, en l'occurrence la variance polygénique résiduelle. Il faut considérer que malgré les efforts dans l'identification de polymorphismes pour des caractères de première importance en sélection, cette partie polygénique *résiduelle* reste encore et dans la plupart des cas la plus importante. Il est important de rappeler ici brièvement le débat suscité il y a déjà deux décennies autour de l'excès de consanguinité apporté par l'implémentation du BLUP (voir par exemple [4]), source aussi d'une augmentation de la précision de sélection via l'évaluation intégrée de pedigrees. Cet enjeu a été à l'origine de la sélection optimale développée subséquemment, et dont on vient de parler dans les pages précédents.

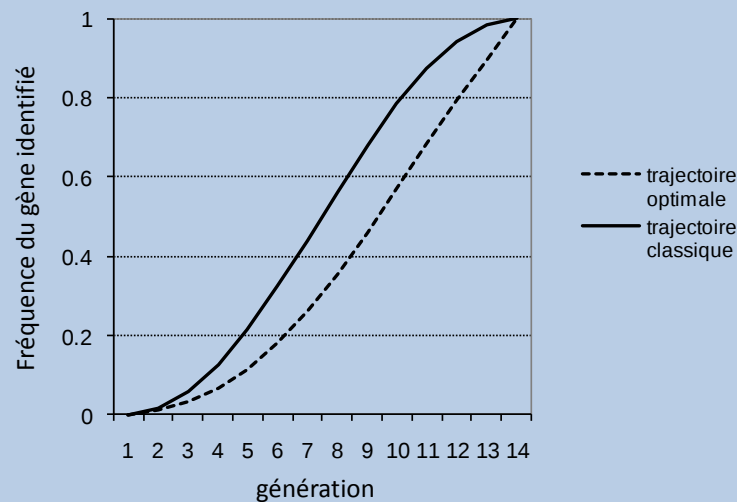
Dans un travail avec Armando Caballero (Université de Vigo, Espagne) et Enrique Santiago (Université de Oviedo, Espagne) ([50], [49]), j'ai abordé la problématique de l'impact de la sélection des gènes

4. Gènes identifiés *versus* fond polygénique: y a-t-il un compromis?

Une trajectoire sélective est constituée de pas discrets de changement de fréquences pour un gène marqué, dont chaque changement est caractérisé par une intensité de sélection (i) au gène :

$$i = \Delta q / [q(1-q)],$$

où le numérateur est un différentiel de sélection sur la forme de changement de fréquences pour ce gène entre générations successives (Δq), et le dénominateur est la variance de fréquences alléliques au gène.



Imaginons une population sélectionnée sur plusieurs générations en gardant une taille constante, un gène majeur favorisé par cette sélection suit une trajectoire typiquement sigmoïde jusqu'à sa fixation (voir **trajectoire classique** dans la figure). Cette courbe est symétrique autour du point où Δq est maximal, avec un nombre équivalent de générations avant et après ce point central. Une alternative optimale à cette trajectoire sigmoïde résulte d'un compromis entre deux termes fondamentalement opposés : le gain génétique au gène identifié, et la taille efficace comme fonction de la variation des contributions génétiques entre candidats en nombre de copies du gène identifié. Suivant des approches quantitatives classiques, on peut décrire ce gain génétique et la taille efficace en fonction de la même variable i , ce qui nous permet de trouver une trajectoire qui maximise simultanément les deux termes (**trajectoire optimale**). Cette trajectoire optimale est celle qui minimise la moyenne des carrés des intensités de sélection, depuis l'apparition du gène d'intérêt dans la population jusqu'à sa fixation. Comparée à la trajectoire classique, la trajectoire optimale ainsi dérivée est asymétrique autour de la génération centrale. Ainsi, la fréquence du gène d'intérêt s'accroît plus doucement que dans la courbe classique pendant les $\frac{2}{3}$ initiaux de la trajectoire, et s'accélère vers le dernier $\frac{1}{3}$ avant la fixation. Une telle asymétrie présente des avantages pour la variation polygénique non identifiée, car ces gènes reçoivent comparativement plus de pression de sélection pendant plus longtemps. Simultanément, ces gènes sont moins susceptibles au déséquilibre de liaison avec le gène d'intérêt. En théorie le LD est plus forte pendant les premiers cycles de sélection (Bulmer 1980) ([11]).

identifiés sur la variation polygénique non identifiée. Effectivement, le processus de fixation par sélection directionnelle des gènes identifiés peut avoir un impact considérable sur le gain génétique provenant des loci non identifiés, ce qu'on connaît sous le nom de **balayage sélectif**. Deux facteurs contribuent au balayage sélectif des loci non identifiés : la réduction de leur intensité de sélection, le poids étant plutôt sur les loci identifiés, et la perte de leur diversité par dérive génétique. Ce phénomène est d'ailleurs bien connu dans le contexte évolutif, et utilisé pour la détection des

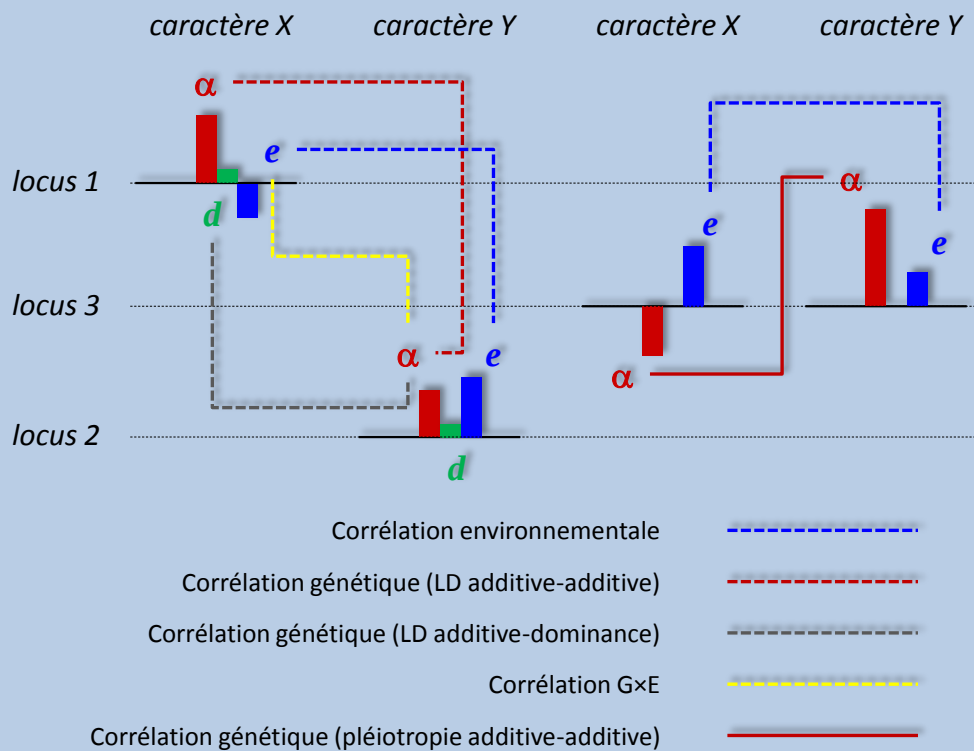
événements sélectifs du passé (par exemple, [12]). Nous avons développé une procédure qui minimise l'effet du balayage sélectif sur la variation polygénique non identifiée considérant une trajectoire sélective optimale pour les gènes identifiés. Le principe est décrit et illustré avec un exemple dans la **boîte 4**. Cette stratégie peut sembler prometteuse dans le contexte de la SAM ou la SAG. Cependant, plusieurs enjeux majeurs restent encore dans l'air, par exemple, comment optimiser plusieurs trajectoires chevauchantes simultanément, ou encore comment combiner ces trajectoires avec d'autres contraintes liées à la gestion de la variabilité dans les populations. D'autres stratégies ont aussi traité la problématique du balayage sélectif en SAM, mais sur une optique très différente basée sur l'utilisation de poids techniques pour les parties identifiées et non identifiées de la variation génétique (voir [72]). Bien que ces stratégies puissent se montrer très flexibles en prenant plusieurs gènes marqués simultanément, il manque une justification théorique explicite pour le choix des poids techniques. Une réflexion globale prenant en compte ces alternatives reste encore à faire. Toutefois, les possibilités croissantes offertes par la génomique au niveau du génome entier peuvent bouleverser ces approches ciblées basées sur quelques marqueurs explicatifs. Certes, les approches génome entier sont encore la *niche* de quelques espèces qui concentrent des ressources génomiques considérables, mais l'accès à une multitude de marqueurs génétiques est en train de devenir possible même pour l'espèce non modèle. Certains questionnements restent cependant toujours d'actualité avec les approches génome entier, comme la gestion de la diversité.

Avec les travaux précédents, nous avons franchi un seuil clé, celui qui sépare les modèles de variation polygénique classiques, comme le modèle infinitésimal, et les modèles dits à hérédité mixte : avec une partie oligogénique ou polygénique connue et une partie polygénique résiduelle non identifiée. Ces derniers modèles hybrides permettent une vision plus explicite de la complexité génique sous-jacente à la variation observée, en vue d'une meilleure compréhension des enjeux en sélection. Nous venons de voir au fil de plusieurs travaux l'enjeu de la perte de diversité après sélection. D'autres enjeux sont liés à la façon dont les caractères sélectionnés partagent la diversité résiduelle après sélection, et plus explicitement aux covariations liant les différents caractères. Les modèles de variation polygénique classiques ne prennent pas en compte d'une façon convenable la covariation entre caractères. Dans les paragraphes suivants, nous verrons comment les modèles à base polygénique connue nous aident à mieux comprendre comment la covariation entre caractères peut avoir un effet sur la sélection.

Des gènes qui s'additionnent, des gènes qui s'opposent

On peut faire le constat suivant : les corrélations entre caractères ont le don de l'ubiquité dans la nature, elles sont universelles et bâtissent les lois de l'allométrie. Toutefois, une partie des ces corrélations est vraisemblablement de nature environnementale. Ceci peut être une explication : les processus ontogéniques sous-jacents aux caractères corrélés peuvent ainsi être affectés par un ou plusieurs facteurs environnementaux communs. Sans doute aussi, la génétique *sensu lato* explique l'autre partie des corrélations observées, et ce sont ces causes génétiques uniquement qui peuvent être les cibles de l'évolution, et donc avoir des conséquences à très long terme dans les populations. Les corrélations génétiques entre caractères, particulièrement les corrélations antagoniques, sont une contrainte majeure dans la dynamique évolutive des populations sélectionnées, et peuvent être le résultat d'une coévolution de caractères simultanément sélectionnés dans le passé ([31]). Mais, qu'est-ce que la nature sous-jacente des corrélations génétiques ? La pléiotropie est une des causes principales de l'existence de corrélations génétiques. Ce phénomène fait appel à des gènes à

5. Mécanismes causals de corrélation phénotypique



Mécanismes sous-jacents de corrélation phénotypique entre deux caractères X et Y pour un système de trois loci : deux ayant des effets uniquement sur un caractère (1 et 2), et le troisième avec des effets pléiotropiques sur X et Y. Ceci n'est pas une représentation exhaustive de toutes les sources possibles dans le système. Les sources de corrélation peuvent être : environnementales entre l'écart e pour locus 1 et celui pour locus 2, ou dans le même locus pléiotropique entre les écarts e pour chaque caractère ; fruit de l'interaction entre effets génétiques et environnementaux ($\alpha \times e$) ; dû au déséquilibre de liaison (LD) entre effets additifs de loci différents ($\alpha \times \alpha$), ou entre effets de dominance et additifs de loci différents ($d \times \alpha$) ; ou dû à la pléiotropie, par exemple, entre effets additifs ($\alpha \times \alpha$).

fonctions multiples, ayant des effets sur plusieurs caractères simultanément ou de façon différée. Une évidence expérimentale toujours croissante (voici quelques exemples : [71], [58], [66], [17]) souligne l'omniprésence des fonctions pléiotropiques dans les génomes. Ce n'est pas difficile d'imaginer l'ubiquité des effets génétiques chevauchants sur plusieurs fonctions, étant donné la complexité et la hiérarchie génétique sous-jacentes au métabolisme des caractères. Un autre facteur à l'origine des corrélations génétiques est le déséquilibre de liaison (LD), sous la forme de covariation entre allèles de loci distincts et affectant différents caractères. Ceci est la conséquence d'un échantillonnage limité par la taille des populations en ségrégation, ou le résultat d'un balayage sélectif. A l'exception faite des loci très proches physiquement dans le génome, le LD est un phénomène transitoire, particulièrement pour les organismes ayant un nombre important de chromosomes ([11]), ou constituant des populations à fort brassage génétique et de grande taille efficace ([32]). Pour la plupart des espèces forestières, ces conditions sont réunies : elles présentent en général plus de chromosomes que d'autres espèces de plantes ([34]), et montrent des niveaux de LD qui décroissent très rapidement et sur des distances courtes le long du génome ([44]). La **boîte 5**

représente graphiquement certains des mécanismes sous-jacents à une corrélation phénotypique entre deux caractères. Pour résumer, bien que la pléiotropie puisse être considérée comme un mécanisme universel, elle n'est pas exclusive et pourrait interagir de façon complexe avec le LD dans des circonstances particulières, notamment en sélection.

Logiquement, les mêmes mécanismes opèrent dans les populations soumises à la sélection artificielle. Assez souvent, l'amélioration génétique comprend plusieurs caractères simultanément, dont certains peuvent être défavorablement corrélés entre eux. C'est le cas assez fréquent en amélioration forestière, de l'antagonisme entre la croissance en volume des arbres et la densité du bois ([80]). La stratégie couramment utilisée dans des programmes de sélection avancés consiste à quantifier des poids économiques pour chaque caractère concerné, et de les intégrer dans un index de sélection multi-caractère. Ces poids doivent être réévalués à chaque cycle de sélection en concordance avec l'évolution des paramètres économiques et génétiques. Aucune supposition n'est faite sur la nature sous-jacente des corrélations, ni sur les effets potentiels de cette nature sur la sélection multi-caractère. Dans le cas des espèces domestiques *historiques*, un grand nombre de générations a pu créer des races caractérisées par des complexes de caractères *coadaptés*, comprenant des valeurs optimales pour chaque caractère, et avec un niveau de corrélation entre caractères peu variable. Contrairement à ces espèces *historiques*, les programmes d'amélioration forestière sont encore dans une phase préliminaire de domestication, dont les effets d'une corrélation adverse entre caractères d'intérêt pourraient ne pas être encore détectables. Cependant, en dépit des enjeux économiques, très peu d'études ont abordé la problématique de la sélection en conditions de corrélations génétiques défavorables. Quelles sont les conséquences de sélectionner deux caractères corrélés défavorablement ? Y a-t-il des stratégies de sélection pour minimiser les effets de cette sélection ?

Partant d'un modèle polygénique à effets connus, j'ai initié une série d'études ayant comme objectif de répondre à ces questionnements, au moins partiellement. Bien que le choix du modèle soit justifié ici, car seulement ces modèles explicites permettent de simuler convenablement des interactions entre allèles et d'assigner des effets de pléiotropie, ces résultats risquent de présenter un manque de généralité. C'est le prix à payer pour construire un modèle explicite. En effet, les connaissances sur la nature des corrélations génétiques sont encore éparpillées, souvent se limitant dans nos espèces d'étude à des estimations quantitatives, et ne permettant pas la paramétrisation complète de ce type de modèles explicites. Cet inconvénient a été contourné ici d'une façon novatrice, en utilisant une approche d'optimisation heuristique. Avec cette procédure, un niveau de corrélation génétique entre caractères est fixé *a priori* comme l'objectif de l'optimisation, et une multitude d'arrangements génétiques caractérisant le modèle à la base peuvent ainsi être évalués selon cet *a priori*. Cela permet une généralisation plus importante, en considérant différents modèles vraisemblables par rapport à l'objectif initial, et facilite la construction de modèles à large nombre de paramètres. Dans la **boîte 6** je représente un exemple très simplifié de modèle polygénique, mais qui demande déjà un nombre considérable de variables. Ces modèles ont permis d'étudier un éventail de conditions où les corrélations génétiques étaient le résultat d'un assortiment variable de LD et de pléiotropie ([52]). Plusieurs conclusions issues de cette publication et d'autres travaux en cours sur les corrélations génétiques peuvent être soulignées ici dans les points suivants :

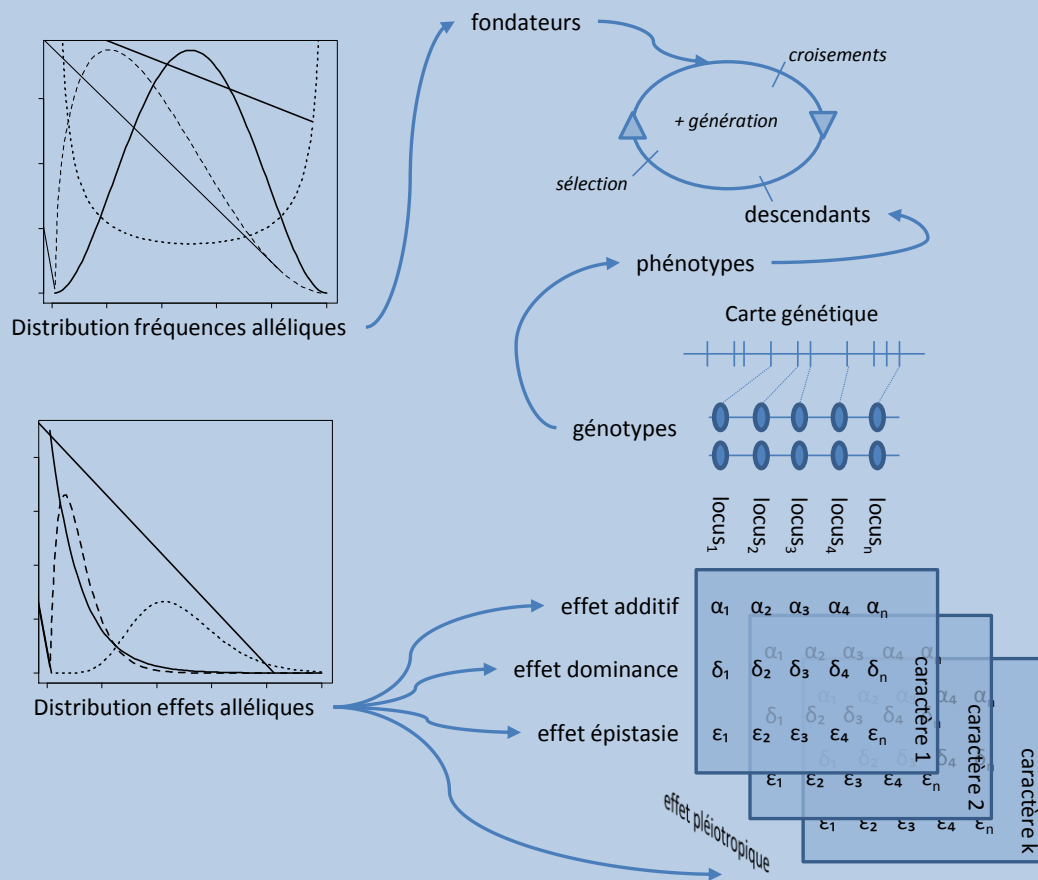
- D'abord, une corrélation génétique forte ou, au contraire, l'absence d'une corrélation observable, ne sont pas des indicateurs des mécanismes génétiques sous-jacents particuliers.

Ainsi, un même modèle comprenant des loci pléiotropiques à effets antagonistes peut donner un éventail très variable de corrélations génétiques, voire des valeurs proches de zéro. Toutefois, les mécanismes pléiotropiques sont en général la cause des corrélations permanentes au fil des générations, avec une tendance vers l'augmentation des corrélations si la sélection va à l'encontre de deux caractères antagonistes. Ce même comportement, cependant, a été observé sur quelques générations dans certains modèles comprenant des caractères issus d'une même région chromosomique à taille très restreinte.

- L'existence d'un mécanisme pléiotropique défavorable à l'évolution conjointe de deux caractères a comme conséquence primordiale une forte augmentation de la variation de la réponse à la sélection pour les caractères impliqués, cet-à-dire, plus de risque associé à la projection du gain génétique. Cette variation, qui résulte de la faillibilité de la sélection pour résoudre l'antagonisme du mécanisme pléiotropique, est d'autant plus forte que la précision de sélection s'améliore via l'évaluation génétique, par exemple, avec l'utilisation d'index optimaux. Même si ce risque est toujours présent dans la praxis, il n'est pas toujours détectable dans des programmes d'amélioration, dans la mesure où la plupart comprennent une seule expérience de sélection.
- Une autre conséquence de la pléiotropie antagoniste en sélection, celle-ci favorable, est la préservation de la diversité génétique neutre et sélective. En effet, l'étude de simulation montre que l'existence d'une corrélation adverse n'entraîne pas une perte additionnelle de variation, en dépit de la faillibilité de la sélection, comparé à un système équivalent sans pléiotropie. Dans le contexte évolutif, la pléiotropie a été déjà proposée comme un des mécanismes impliqués dans la conservation de la diversité adaptative ([40]).
- Finalement, peut-on identifier une stratégie qui permet une réduction des effets défavorables issus de corrélations génétiques ? Cette étude a permis de révéler quelques indices de réponse. Ainsi, la comparaison de deux stratégies classiques très contrastées, la sélection basée sur un index optimal multi-caractère et la sélection par niveaux indépendants, a permis de mieux comprendre les mécanismes agissant sur la génération de la variance du gain génétique. Toujours dans le cadre réductionniste des modèles abordés, dont l'additivité des effets est majoritaire sur d'autres effets non additifs comme la dominance, les régimes de sélection qui préservent plus efficacement l'hétérozygotie sont aussi ceux qui réduisent les effets défavorables liés aux corrélations pléiotropiques. Ceci a été montré théoriquement au cours de mes simulations en implémentant une sélection optimale, comme celles décrites précédemment, avec une contrainte explicite sur la perte de diversité génétique.

Pour conclure, en répondant aux deux questions qui ouvrent cette section, [1] la sélection de caractères corrélés défavorablement devient plus faillible, moins précise, que dans le cas d'absence de pléiotropie ou de LD, avec un risque accentué de perte de réponse pour les caractères impliqués comme l'effet le plus marquant ; et [2] l'alternative suggérée ici implique les mêmes outils et objectifs que ceux préconisés pour le contrôle de la dérive génétique. Il faut rappeler que dans le cadre d'une sélection mono-caractère, le cas le plus amplement traité dans la littérature sur la sélection artificielle, l'augmentation de la taille efficace a déjà été identifiée comme un élément clé dans la réduction de la variance de la réponse (voir le travail classique [30]). Le point [2] confirme ainsi dans un contexte plus général la théorie établie, pour laquelle la variation des fréquences des allèles par dérive est à l'origine de la variation de la réponse en sélection. Toutefois, l'apport novateur de mes travaux concerne la prise en compte d'un système multi-caractère dont les

6. Modèle allélique explicite: une multiplicité de variables



Exemple simplifié d'un modèle allélique explicite, qui montre certaines des variables à prendre en considération pour la construction d'un génotype. Le processus commence avec l'échantillonnage dans une distribution de fréquences alléliques pour bâtir le pool génétique des fondateurs. Dans les générations successives, chaque nouveau descendant recevra par descendance des allèles, dont les effets sont spécifiés a priori à partir d'une distribution.

corrélations répondent à des mécanismes sous-jacents vraisemblables, et à une meilleure compréhension des enjeux en sélection via l'utilisation des modèles polygéniques à effets connus.

L'évolution par sélection naturelle n'est qu'un processus d'optimisation multifactoriel, mais extrêmement puissant. Ceci a inspiré le développement d'algorithmes d'optimisation presque aussi efficaces, comme les algorithmes génétiques. La sélection artificielle n'est qu'un modèle à petite échelle de cette même évolution et, jusqu'ici, j'ai essayé de la présenter comme un processus d'optimisation sous différentes contraintes. L'objectif dans les programmes d'amélioration avancés est ainsi d'implémenter une gestion raisonnée de la diversité génétique disponible en sélection pour garantir une réponse à long terme. Toutefois, le concept de diversité sur lequel on repose ici considère que les différences entre individus peuvent être expliquées *in fine* par la substitution d'allèles dans leurs génotypes, ce qui correspond à une extension du paradigme additif en génétique ([22]). Bien que ce concept reste valide dans le domaine de la prédiction, on néglige le fait que derrière chaque différence observable il y a un historique de succession d'effets, certains en

interaction, qui créent un développement particulier, une trajectoire. Ce sont des altérations dans ces programmes de développement individuel qui sont fondamentalement la cible de la sélection naturelle (voir par exemple [39]). Dans ce sens, la plasticité phénotypique nous donne un exemple significatif et souvent mesurable à courte échelle du développement en action face à des milieux hétérogènes. Dans la section suivante, j'introduirai brièvement les concepts autour de la plasticité phénotypique, le pourquoi de son intérêt et les recherches dans lesquelles je suis impliqué au sein de mon groupe de collaborateurs.

L'étude de la plasticité phénotypique

Qu'est ce que la plasticité phénotypique ? La littérature scientifique nous offre une multitude de définitions plus ou moins laxistes concernant le contour du phénomène. J'ai élaboré ici une définition restrictive, pour laquelle la plasticité phénotypique est **un changement associé à un ou plusieurs caractères qui résulte de l'exposition d'un même génotype à une variation environnementale**. Elle est restrictive dans le sens où le sujet du changement est le génotype comme individu, par rapport à d'autres acceptions moins strictes de génotype, et étrangères à celle de la génétique, comme population ou encore l'espèce. Elle implique aussi des états phénotypiques multiples par génotype, successifs dans le temps ou simultanés dans l'espace, en opposition aux mesures ponctuelles qui caractérisent généralement un caractère quantitatif quelconque. La plasticité phénotypique est de ce fait une trajectoire individuelle pour un caractère dans un environnement hétérogène. Cette trajectoire peut être modélisée statistiquement en fonction des variables environnementales qui créent l'hétérogénéité du milieu, et on appelle **norme de réaction** (NdR) la fonction qui en résulte. L'interaction génotype $\times$ environnement n'est qu'une expression de la variation de NdR dans une population. Une NdR peut être construite à partir de réponses de différentes copies d'un même génotype installées dans des milieux distincts, comme c'est souvent le cas des dispositifs expérimentaux (**plasticité spatiale**) ; ou comme une véritable trajectoire individuelle dans le temps dans un milieu hétérogène (**plasticité temporelle**). Dans le premier cas, la courbe résultante présente la particularité statistiquement désirable d'être issue d'expériences indépendantes, mais le choix arbitraire des milieux risque de fausser la pertinence biologique et évolutive de la courbe. Dans le cas de la plasticité temporelle, le risque vient de l'autocorrélation entre points successifs, mais la réponse résultante peut avoir une pertinence adaptative directe.

Le changement environnemental global, avec sa composante climatique prépondérante, est incontestablement un challenge majeur pour des nombreux écosystèmes et leurs espèces constituantes. Ce challenge pourrait se révéler spécialement difficile pour les espèces longévives comme les arbres, qui doivent supporter à maturité des conditions très différentes de celles de leur installation ([25]). Toutefois, le caractère immobile et la longévité des arbres ont pu façonner évolutivement leur capacité plastique, d'une part pour élargir leur éventail d'habitats potentiels et, d'autre part pour créer un tampon face à l'extinction ([28]). Cette hypothèse trouverait un support considérable dans l'expérience accumulée grâce au vaste réseau de dispositifs expérimentaux de l'amélioration forestière. En effet, certains génotypes montrent une capacité étonnante d'acclimatation à des environnements très contrastés, capacité mesurée sur des caractères très intégrateurs comme leur taux de croissance. Peut-on quantifier cette particularité plastique des arbres d'une façon plus exhaustive, et sur quels caractères ? Existence-ils des différences génétiques pour cette capacité plastique ? Les réponses à ces questions permettraient d'avoir des éléments

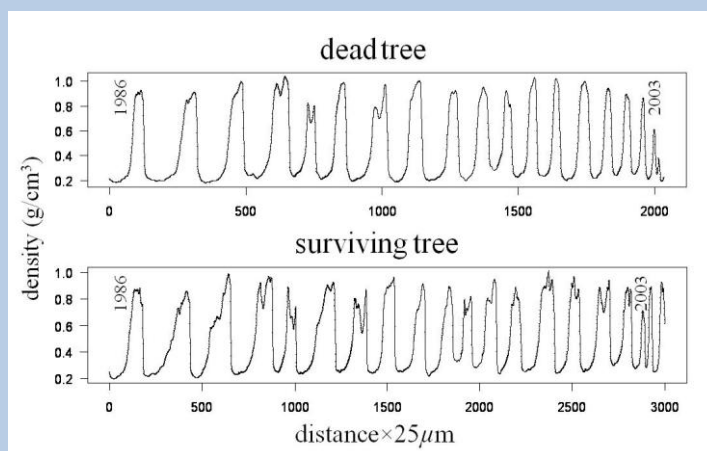
encore plus objectifs pour la gestion raisonnée des populations servant comme source pour la reforestation.

Pour répondre à la question sur les caractères, le bois est un objet d'étude incontournable. Il joue des fonctions de support physique pour l'appareil photosynthétique et reproducteur, et aussi de conduction de la sève entre ces différents composants fonctionnels. Pendant la formation du bois, le cambium des arbres interagit avec les signaux du milieu en adaptant l'anatomie des nouvelles cellules aux changements ainsi aperçus par la plante. Cette modification laisse une trace anatomique permanente dans le bois pour chaque saison de croissance, une trajectoire qui devient un enregistrement pérenne de la réaction du tissu vivant à l'environnement de l'arbre. Le rôle adaptatif serait lié à l'harmonisation entre les propriétés conductrices du bois ainsi formé, les changements dans l'environnement de la plante et ses demandes vitales (voir **boîte 7**). Ainsi, des cellules conductrices à grand diamètre permettraient les forts débits en sève nécessaires pour une croissance rapide, mais elles seraient aussi spécialement susceptibles à l'embolie provoquée par un stress hydrique et source potentielle de dysfonction hydraulique ([16]).

Comment lire cette trajectoire dans le bois ? Brièvement, la radiographie à rayons X d'un échantillon radial de bois, depuis le cœur de l'arbre jusqu'à l'écorce, permet la lecture rétrospective des modifications anatomiques. Ces modifications sont ainsi transformées via la radiographie en profils de densité, nommés de *microdensité*, et dont la trajectoire laisse des limites bien reconnaissables dans les cernes annuels. Historiquement, la dendrochronologie a utilisé cette approche pour la datation et l'inférence climatique du passé, en se basant sur une réponse moyenne des arbres. Toujours sur le principe de lecture rétrospective du bois, Philippe Rozenberg et collaborateurs ont ouvert un champ de recherches nouveau, dont l'objet d'étude est la variation individuelle et entre individus dans leurs profils de microdensité comme réaction à des événements climatiques particuliers, voire extrêmes. Le néologisme **dendroplasticité** a été adopté par le groupe pour décrire d'une façon générale la NdR temporelle contenue dans le profil de microdensité, et assimilée à une **NdR développementale** (selon [59]). En effet, la lecture rétrospective de la dendroplasticité dans des peuplements soumis à des contraintes environnementales nous offrirait la possibilité de mieux quantifier les différences adaptatives entre les individus et *in fine* de comprendre un des mécanismes sous-jacents à l'adaptation des arbres.

Depuis la naissance du concept dendroplasticité, je suis impliqué en tant que généticien dans ces recherches, notamment sur l'étude de la dendroplasticité comme caractère quantitatif et sur l'évaluation des composantes héréditaires et adaptatives dans les profils issus de populations artificielles et naturelles. Un premier effort a été porté sur la construction d'une NdR dendroplastique fondée sur des index climatiques simples, comprenant des variables de température et de précipitation propres ou proches des populations mesurées. Notre premier exemple comporte un dispositif clonal de Pin maritime sur 3 sites [55], où nous avons estimé individuellement les pentes d'une NdR temporelle linéaire liant les changements en microdensité aux changements dans l'index climatique, pour une année spécifique avec des sécheresses ponctuelles. L'héritabilité *sensu lato* des NdR ainsi obtenues montre des valeurs modérées et des niveaux de variation génotypique pour la pente relativement importantes. Les mêmes principes de construction de NdR ont été repris et perfectionnés dans la thèse d'Alejandro Martínez-Meier (2009) (directeurs Mario Pastorino de l'INTA en Argentine et Philippe Rozenberg), et dont j'ai participé à l'encadrement. Cette thèse, focalisée sur la réponse du Douglas à la canicule de l'année 2003, nous montre sur plusieurs dispositifs

7. Dendroplasticité: une norme de réaction dans le bois



Deux profils de microdensité au fil des cernes, pour un arbre mort (supérieur) après la canicule de l'année 2003, et son voisin toujours vivant à l'époque de la récolte (2005). La distance dans l'axe des abscisses correspond au rayon depuis le cœur de l'arbre (gauche) jusqu'à l'écorce (droite). Sur l'axe des ordonnées, la valeur en densité du bois qui résume les changements anatomiques le long de la saison de croissance de l'arbre. Chaque pic de densité correspond au bois produit pendant la saison sèche (bois final), et les creux en densité correspondent au bois moins dense produit pendant la saison humide (bois initial). Les arbres morts dans le dispositif auquel correspondent ces deux échantillons présentent en général un bois moins dense et des cernes comportant très peu de bois final. Depuis Martinez-Meier et al (2008) ([37]).

expérimentaux que cet événement extrême a affecté considérablement la formation du cerne de cette année, avec une croissance exceptionnellement faible, voire interrompue pour certains individus. Toutefois, les arbres mesurés présentent des réactions dendroplastiques différentes les uns aux autres, et une partie importante de ces différences en NdR observées est d'origine génétique ([38], [36]).

Bien que les valeurs d'héritabilité modérée trouvées pour les variables décrivant les NdR suggèrent déjà un potentiel de réponse valorisable en sélection artificielle ou, éventuellement, par la sélection naturelle, il faudrait encore estimer par expérimentation la valeur adaptative de cette variabilité et, éventuellement, identifier des NdR favorisées sous certaines contraintes. L'événement exceptionnel de la canicule a provoqué de nombreux phénomènes de dépérissement et de mortalité dans les populations de l'espèce Douglas, à juste titre une opportunité rare pour observer la sélection naturelle en action. Dans un autre chapitre de la thèse de Martínez-Meier (voir [37]), nous abordons l'étude des différences systématiques en dendroplasticité entre individus morts après la canicule et les voisins le plus proches ayant survécu, se trouvant dans un dispositif âgé de comparaison de provenances de Douglas. Les résultats montrent que certaines variables décrivant la NdR dendroplastique, et identifiées dans les études précédentes comme ayant des héritabilités modérées, se montrent significativement différentes entre individus morts et leurs voisins vivants, et cette différence est systématique sur plusieurs années avant la canicule. Brièvement, les individus survivants présentent des cernes à plus forte densité de bois, et cela suggère une plus haute

résistance à l'embolie en conditions extrêmes de canicule (**boîte 7**). Toutefois, on est conscient que ce résultat est corrélationnel, et ne permet pas une inférence causale au delà de l'hypothèse déjà évoquée de l'embolie. Ainsi, nous devons encore établir le lien causal entre maladaptation et embolisme, et entre ce dernier et la dendroplasticité, en identifiant des NdR favorables et défavorables.

De nouveaux efforts ont été portés sur une gamme plus ample de dispositifs expérimentaux de l'amélioration génétique, et sur des populations faiblement anthropisées. Dans le cadre des expériences issues de l'amélioration, je dois mettre en évidence ici deux thèses en cours dont je participe à l'encadrement. L'une plus focalisée sur les aspects écologiques autour de l'expression de la dendroplasticité, et l'autre concernant des aspects de génétique quantitative. La première thèse (étudiante Anne-Sophie Sergent, directeurs Nathalie Bréda de l'INRA Nancy et Philippe Rozenberg) concerne l'étude du dépérissement du Douglas sur un vaste échantillonnage de dispositifs de comparaison de provenances couvrant deux des régions les plus affectées par la canicule de l'année 2003 (Bourgogne et Midi-Pyrénées). L'objectif de la thèse est d'identifier les conditions climatiques et pédologiques, sylvicoles et biotiques du dépérissement dans ces deux régions, et de bâtir un lien avec la dendroplasticité, en vue de caractériser des NdR susceptibles d'être associées à une maladaptation future. Dans la deuxième thèse (étudiante Manuela Ruiz-Diaz, directeurs Luis Mroginski de l'Université *Nacional del Nordeste* en Argentine et Philippe Rozenberg), une analyse est portée sur plusieurs dispositifs de descendance maternelles, une partie ayant fourni les candidats à la sélection pour le programme d'amélioration Douglas. L'objectif de la thèse consiste à fournir des estimations d'héritabilité *stricto sensu* pour la NdR dendroplastique et ses composants, jusqu'à maintenant caractérisées uniquement par des estimations *sensu lato*; et de quantifier les corrélations génétiques additives entre dendroplasticité et d'autres caractères de croissance et phénologiques faisant partie des critères du programme d'amélioration. Ces deux thèses combleront le manque principal dans nos travaux précédents, notamment les tailles relativement réduites des expériences, en vue d'une meilleure généralisation des résultats.

J'ai aussi évoqué des travaux sur des populations faiblement anthropisées comme alternative aux dispositifs expérimentaux de l'amélioration génétique. Les dispositifs artificiels étudiés jusqu'ici sont souvent limités par rapport à la gamme d'environnements couverts. Parmi les populations faiblement anthropisées, celles issues des gradients altitudinaux de montagne représentent un modèle exceptionnel, car les environnements parcourus le long de distances relativement courtes sont souvent extrêmes, et abritent des écosystèmes particulièrement menacés par le changement climatique. Dans le cadre de plusieurs projets en cours coordonnés par Philippe Rozenberg, nous caractérisons dans le continuum populationnel d'un gradient altitudinal de mélèze alpin la dendroplasticité temporelle et spatiale présente dans un échantillonnage intensif de plusieurs milliers d'individus distribués sur plusieurs niveaux d'altitude, et nous sommes engagés au sein du même dispositif à initier une expérience de transplantation croisée couvrant le gradient. L'objectif général est d'obtenir des éléments de réponse sur la façon dont la plasticité est imbriquée à l'adaptation locale. Dans cet objectif, mon engagement principal est l'implémentation d'outils de génétique quantitative permettant l'estimation des paramètres génétiques sans connaissance *a priori* des relations d'apparement, et via l'utilisation de marqueurs moléculaires ([5]). Toutefois, la validité de cette méthode d'estimation est très dépendante des propriétés génétiques et démographiques de la population d'étude, et reste sensible aux interactions génotype-

environnement. D'autres travaux sont en cours dans notre équipe pour déplacer ces limites, notamment avec le concours de la statistique spatiale.

On a eu un aperçu de la dendroplasticité comme modèle pour l'étude de la plasticité adaptative, et dont les premières étapes concernant l'implémentation de la génétique quantitative sont déjà bien engagées. Ce même objet d'étude, la formation du bois au fil de la croissance de l'arbre, est abordé au sein de mon unité par des approches et des équipes complémentaires, développant notamment la transcriptomique, une quantification systématique des ARN messagers pour inférer l'importance relative des différents gènes actifs dans le bois vivant. In fine, ces approches complémentaires visant d'un côté le phénotype et d'autre côté les gènes, devront se trouver harmonisées dans un modèle architectural et fonctionnel joignant les deux extrêmes. Dans la section suivante, je présente un projet de recherches visant cet objectif.

3. Perspectives : projets de recherche engagés et futurs

Une preuve indirecte de la complexité des génomes vient de la génomique comparée entre espèces phylogénétiquement proches. Prenons un exemple bien proche de nous : celui des espèces d'hominidés. En dépit des différences phénotypiques (considérables) qui semblent nous séparer, nous les êtres humains des autres hominidés, chimpanzés et gorilles, les différences génomiques sont étonnamment petites : entre 2 et 3% ([43]). Ceci, semble être le cas d'autres binômes d'espèces, et même pour celles moins proches phylogénétiquement. Un exemple serait la comparaison entre le génome du peuplier ([70]) et celui de plantes annuelles comme *Arabidopsis*, qui révèle aussi des différences relativement petites entre ces taxons distants en relation aux différences *visibles* [24]. Une des interprétations possibles de pourquoi « *aussi peu détermine autant* » de différences viendrait de l'architecture hiérarchique du génome basée sur l'ubiquité des réseaux de régulation et d'interaction entre gènes. Cette interprétation, faisant appel aux interactions entre gènes, est loin du paradigme opérationnel qui a marqué autant la génétique jusqu'à nos jours, et défini par une multiplicité de gènes fondamentalement indépendants à fonction unique et petit effet.

Cette opposition conceptuelle sur la façon de modéliser le phénotype à partir du génotype est déjà présente au moment de la genèse de la synthèse évolutive moderne, et énergétiquement alimentée par deux de ces piliers : Sir Ronald Fisher et Sewall Wright. Fisher ([22]) conçoit l'architecture des caractères selon un modèle additif, où les gènes opèrent comme des facteurs indépendants à effets virtuellement infinitésimaux : le **modèle infinitésimal**. Cette conception a permis la fondation de la génétique quantitative par le rapprochement *de facto* entre génétique et statistique : la variation continue des caractères observée par les statisticiens peut découler facilement du modèle Fishérien. Comme alternative à cette vision statistique et populationnelle, Wright propose une vision mécaniste et centrée sur l'individu ([78]), où l'architecture génétique est basée sur l'ubiquité des gènes à effets multiples (**pléiotropie**) et l'interaction entre gènes (**épistasie**). Bien que l'interprétation Fishérienne ait épaulé avec succès la théorie de la sélection artificielle et de l'amélioration génétique ([21]), les mécanismes sous-jacents à l'évolution du phénotype ne peuvent pas être traités conformément aux mêmes principes statistiques additifs. Pour Fisher, par exemple, l'épistasie est perçue d'une façon uniquement statistique, comme une source de variation non additive. Toutefois, Wright avait négligé le rôle d'un environnement changeant, en accentuant uniquement l'importance des mécanismes génétiques dans l'expression du phénotype.

Bien que le modèle Fishérien reste un choix opérationnel pour la prédiction de l'évolution phénotypique, la conception de Wright est considérée de nos jours comme le modèle le plus vraisemblable d'architecture génétique des caractères. Il y a des évidences multiples d'interaction entre gènes pour les organismes unicellulaires, comme *Escherichia coli* ou *Saccharomyces cerevisiae*, et pour certains organismes pluricellulaires modèles (*Drosophila* et le nématode *C. elegans*), où de vastes collections de délétions génomiques ponctuelles ont permis de cartographier finement des réseaux de gènes ([20]). Des études d'expression des gènes, via la concentration d'ARN messenger, ont révélé dans les mêmes modèles précédents que ces réseaux obéissent, dans leur géométrie et niveau d'interactions, aux changements physiques du milieu ([35], [79]). Les différences au niveau de l'expression génétique induites par le milieu sont vraisemblablement le résultat d'une conception modulaire de l'architecture génétique des caractères, où des réseaux de gènes sont organisés en sous-réseaux répondant à des signaux internes ou externes spécifiques ([61]). En biologie, les réseaux (protéine-protéine, gène-gène, entre espèces) suivent une norme *universelle*, selon laquelle la

plupart des nœuds se caractérisent par très peu d'interactions, hormis quelques nœuds centraux qui ont des interactions multiples ([3]). La particularité de cette topologie (« *scale-free* ») est un très court acheminement d'interactions entre nœuds extrêmes, appelé diamètre du réseau, ce qui facilite des réponses rapides aux changements du milieu ([60]).

La pléiotropie et l'épistasie constituent les phénomènes interactifs de base dans les réseaux de gènes, avec une évidence expérimentale considérable ([71]). Le concept de pléiotropie a été déjà présenté plus tôt, lié aux causes des corrélations génétiques. Dans les réseaux de gènes, le même concept est toujours valable : celui d'un gène affectant plusieurs modules fonctionnels, ou sous-réseaux, chacun attaché à l'expression d'un caractère différent. L'autre phénomène, l'épistasie, résulte du masquage ou de la modulation de l'effet d'un gène par un autre gène, et peut se manifester notamment dans la régulation génétique et la transduction de signaux internes ou externes. Ces deux phénomènes s'imbriquent profondément dans l'architecture du réseau pour révéler ou masquer la *variation* génétique sous-jacente (celle des polymorphismes), en montrant la *variabilité* génétique (celle des combinaisons génotypiques) et phénotypique sur lesquelles agissent *in fine* les forces évolutives.

Le projet que je présente ici a pour objectif général **d'étudier d'un point de vue mécaniste par modélisation le lien entre gène et phénotype**, en prenant en compte l'architecture en réseau des génomes et l'expression développementale des phénotypes en NdR. Cet intérêt est bien enraciné dans mon expérience en modélisation Fishérienne, dont nous avons eu un aperçu dans ce mémoire ; largement motivé par le challenge explicatif récent autour de la plasticité ; et *in fine* utile pour la prise en compte d'autres sources de variation dans le processus d'évaluation et sélection artificielle. Hormis quelques exceptions très spécialisées (voir [27]), le manque de modèles équivalents dans la littérature scientifique peut justifier cet investissement. Cet objectif général est décliné en questions scientifiques plus détaillées dans les sections suivantes.

La plasticité phénotypique comme expression des interactions gène-gène et gène-environnement

Comment l'architecture des caractères (degré d'épistasie, de pléiotropie et leur imbrication) peut affecter l'expression de la plasticité phénotypique ? Comment les différents types de topologies de réseaux de gènes réagissent à l'imprédictibilité de l'environnement ? Quels régimes de sélection (intensité, type de troncation) et d'hétérogénéité de l'environnement (niveau et degré de prédictibilité) facilitent la préservation de gènes impliqués dans l'expression de la plasticité phénotypique ? Ces questions ont motivé un projet de développement d'un outil de simulation spécifique et hautement paramétrable. Ce développement s'insère dans un projet collaboratif à l'échelle européenne, NOVELTREE (2008-2012), et comporte un poste postdoctoral du Ministère de la Recherche (2009-2011), qui a permis sous ma direction le codage et la mise au point d'un simulateur par une ingénieure informatique, spécialisée auparavant dans les réseaux cellulaires. L'utilisation de cet outil pour répondre à une partie de ces questions pourrait faire en suite l'objet d'une thèse, dont le financement reste encore à trouver.

Dans cet outil de simulation, le contrôle génétique de la réponse plastique comporte plusieurs mécanismes putatifs paramétrables :

- **L'hétérozygotie**, mécanisme selon lequel des allèles distincts confèrent des réponses spécialisées, et donc couvrant en hétérozygotie un plus large spectre d'activité, est considérés le

moins plausible des mécanismes sur la base de l'absence d'évidences expérimentales sur des espèces modèles.

- La **sensibilité allélique**, mécanisme par lequel les gènes adaptent directement leur expression en fonction du milieu. Des exemples abondent concernant la dépendance entre activité enzymatique et température, pH ou la concentration des substrats. Souvent, ce mécanisme est associé à une forme de pléiotropie, dans laquelle un gène montre des expressions alternatives en fonction du milieu. C'est un mécanisme passif de réponse plastique.
- La **régulation génétique**, ou l'épistasie proprement dite, mécanisme où des gènes régulateurs sont les cibles des signaux environnementaux. C'est un mécanisme actif de réponse plastique.

Ces mécanismes seront étudiés indépendamment ou en combinaison à l'aide des modèles alléliques structurés en réseau de gènes plus ou moins interdépendants, avec des architectures géniques sous une forme hiérarchique « *scale-free* » (dépendances unidirectionnelles) ou en filet (interdépendances), et à diamètres du réseau variables. La NdR plastique résultante sera une fonction multidimensionnelle de l'expression du génotype face à une variation donnée du milieu. Cette modélisation permettra aussi la création de lignages sur plusieurs générations, avec la possibilité de confronter une estimation de paramètres génétiques pour une NdR « *dégradée* », à partir de groupes d'apparentés, et celle « *idéale* » issue d'un même génotype.

D'une façon équivalente aux modèles de corrélation génétique détaillés plus haut, l'optimisation heuristique sera utilisée pour rendre les résultats plus généralisables, en générant des modèles alléliques en réseau compatibles avec des variables populationnelles connues *a priori*, par exemple, le niveau d'héritabilité ou de variation génétique pour chaque caractère. Ceci permettra de mieux prendre en compte des observations issues de l'expérimentation, notamment les résultats de l'expérience de transplantation croisée couvrant le gradient altitudinal de mélèze alpin (Projet GRAAL, GICC), ou les thèses sur la dendroplasticité en cours. D'autres résultats sont attendus au sein de notre unité concernant la topologie des réseaux impliqués dans la formation du bois. En effet, plusieurs travaux visent à révéler la hiérarchie des gènes sous-jacents à l'activité cambiale avec des approches transcriptomiques ([19], [33]). À court terme, la demande d'un poste de Chargé de recherches sur l'inférence de la topologie des réseaux à partir des données de la transcriptomique permettrait d'avoir une « personne d'interface » entre les études d'expression et la conception et l'implémentation de modèles mécanistes.

Vers une prise en compte des sources d'interaction G×G et G×E dans l'évaluation génétique de nos programmes d'amélioration

Grâce aux modèles polygéniques à effets connus développés précédemment pour l'étude des corrélations génétiques ou le balayage sélectif, et dans le cadre du projet collaboratif européen NOVELTREE, nous développons un outil d'évaluation des performances de la SAM *in silico* pour construire chez le peuplier une résistance durable à la rouille. En effet, rares sont les cas où cette approche sélective a été appliquée en routine pour une espèce végétale, et aucun cas n'existe à notre connaissance pour une espèce forestière en dépit des avantages potentiels associés à cette alternative. L'outil est construit pour intégrer des résultats obtenus *in vivo*, à partir du programme d'amélioration de l'espèce, dans un modèle simplifié d'architecture génomique pour des caractères de résistance à la rouille. Ces résultats comprennent notamment le nombre de QTL ayant un effet sur la résistance, les effets des QTL, le nombre et type de marqueurs et leur position sur une carte

génétique, ainsi que les taux de recombinaison le long du génome et l'estimation des niveaux du LD. En partant d'une situation la plus proche possible de la *réalité*, l'objectif est de prédire les effets de l'évaluation et de la sélection selon différents scénarios pour optimiser *in fine* l'implémentation de la SAM dans la population d'amélioration du peuplier. La performance ne sera qu'un des critères à observer, le niveau de diversité aux marqueurs sera aussi prise en compte objectivement, notamment au niveau intrafamiliale. Un point clé dans ce processus d'optimisation *in silico* sera la mise au point des outils statistiques nécessaires pour l'évaluation des performances individuelles, et allant de la reconstruction d'haplotypes et la matrice d'identité par descendance aux marqueurs, jusqu'à la résolution du modèle mixte pour l'obtention du BLUP individuel. Cette étape de mise au point fera l'objet d'un travail d'ingénieur pour la période 2011-2012, dont je serai un des encadrants principaux.

Les avances rapides dans les techniques de séquençage, rendant possible les approches génome entier, en génomique fonctionnelle et transcriptomique, et le phénotypage haut débit comme le NIRS (« *Near-infrared spectroscopy* »), permettent déjà une caractérisation plus nette du lien gène-phénotype, et éventuellement l'inférence des interactions gène-gène et gène-milieu dans certains cas. Il est envisageable d'alimenter le modèle prédictif pour la SAM avec les approches *réseau de gènes* développées pour l'étude de la plasticité phénotypique, et ainsi permettre une estimation plus propre des composantes additives et non additives pour la performance et la variabilité génotypique au fur et à mesure où ces données d'interaction sont disponibles. Une réflexion est engagée dans le cadre du projet NOVELTREE et avec d'autres partenaires extérieurs sur la faisabilité de la sélection pangénomique pour nos espèces d'étude. Nous sommes encore loin du modèle pour la floraison d'*Arabidopsis*, pour lequel une connaissance fine de la topologie du réseau des gènes impliqués et la prise en compte des signaux environnementaux de lumière et température ont permis une reconstitution virtuelle précise du comportement phénologique ([27]).

Note finale

Je ne voudrais pas oublier d'autres activités qui font partie de mon travail *quotidien* comme généticien et améliorateur. Les deux chantiers de recherche précédents requièrent une bonne connaissance sur les caractères d'intérêt, les programmes existants, et leurs contraintes opérationnelles. Dans ce sens, je me suis engagé dans la gestion du programme d'amélioration du Douglas, où je travaille à la constitution de la nouvelle population d'amélioration, en collaboration notamment avec Jean-Charles Bastien. Cet engagement représente une source inestimable d'apprentissage dans ma récente voie forestière, et des opportunités pour diriger des nouveaux travaux de recherche.

4. Liste de projets de recherche

Dans la liste suivante je détaille les projets où j'étais ou suis impliqué, avec quelques mots clés indiquant la thématique de ma contribution.

- 2003 : Expertise pour l'*INSTITUTO DE INVESTIGAÇÃO DA FLORESTA E PAPEL* (Lisbonne, Portugal) ; utilisation de marqueurs moléculaires pour l'inférence de l'apparentement et le *fingerprinting* pour le programme d'amélioration d'Eucalyptus.
- 2003 : Expertise pour le *British Columbia MINISTRY OF FOREST Research Branch* (Victoria, Canada) ; gestion des corrélations génétiques adverses en sélection.
- 2004 : Projet Innovant du Département **INRA EFPA**, « *Modélisation: un outil d'aide au choix des stratégies optimales, pour la gestion des populations d'amélioration, et la constitution de réseaux de conservation des ressources génétiques* » ; coordination.
- 2004-2006 : Projet de recherche du **BUREAU DES RESSOURCES GENETIQUES**, « *Modélisation et comparaison de la diversité neutre et non neutre dans les collections et les populations naturelles de deux espèces ligneuses: le peuplier noir et le merisier* » (coordonnateur Frédérique Santi, INRA Orléans) ; responsable des activités de simulation.
- 2004-2006 : Projet de recherche espagnol du MINISTERIO DE CIENCIA Y TECNOLOGIA, « Effet de la sélection de gènes spécifiques sur la variation et la structure génétique des populations ». (coordonnateur Armando Caballero, **Université de Vigo**) ; modélisation, génétique de populations.
- 2004-2007 : Projet collaboratif Europe - Amérique du Sud du Programme **Alpha II GEMA** de l'UE, « Génétique du bois » (coordonnateur Philippe Rozenberg) ; encadrant de plusieurs étudiants et post-docs.
- 2005 : Expertise pour le *British Columbia MINISTRY OF FOREST Research Branch* (Victoria, Canada) ; optimisation des schémas de sélection et d'accouplements pour le programme Douglas en BC.
- 2005 : Projet Innovant du Département **INRA EFPA**, « *Neutral markers and relatedness in the estimation of adaptive genetic variability in natural forest* » ; coordinateurs Leopoldo Sanchez et Philippe Rozenberg.
- 2006-2009 : Projet de l'Agence Nationale de la Recherche, **DRYADE**, « *Vulnérabilité des forêts face aux changements climatiques : de l'arbre aux aires bioclimatiques* » (coordonnateur Nathalie Breda, INRA Nancy) ; activités en génétique quantitative et dendroplasticité sur le Douglas.
- 2007-2010 : Projet collaboratif européen du 6ème Programme cadre de l'UE, **TREEBREDEX** (coordonnateur Luc Pâques) ; coorganisateur des activités en réseau et deux ateliers internationaux.
- 2007-2010 : Projet du Ministère de l'Agriculture, **PAF-Dendroplasticité** (coordonnateur Philippe Rozenberg) ; responsable des activités en génétique quantitative et de populations dans le gradient altitudinale Alpin.
- 2007-2010 : Projet Espagnol de recherche, « *Qualité structurale du bois de Pinus pinaster en Galice, variabilité et plasticité phénotypique et méthodes non destructives d'évaluation* » (coordonnateur Esther Merlo, Institut **CIS Madeira**) ; activités en génétique quantitative et dendroplasticité.
- 2008-2011 : Projet bilatéral France-Argentine, **ECOS-SUD ARGENTINE** (coordonnateur Philippe Rozenberg) ; activités en génétique quantitative et dendroplasticité sur le gradient altitudinal Andin.
- 2008-2012 : Projet collaboratif européen du 7ème Programme cadre de l'UE, **NOVELTREE** (coordonnateur Catherine Bastien) ; leader d'activité "Providing strategies to overcome breeding constraints in forest trees".
- 2009 : Sir Frederick McMaster Fellowships, CSIRO (Australie) « *Virtual Genes: development of a locus-based gene system* ».
- 2009-2010 : Projet bilatéral France-Espagne, **PICASSO** (coordonnateur Philippe Rozenberg) ; activités en statistique et génétique quantitative sur la dendroplasticité du pin maritime en Galice.

2009-2011 : Projet de la Région Centre « Empreinte du changement climatique sur le bois des forêts de la Région Centre » (**XYLOME**) (coordonnateur Philippe Rozenberg) ; activités en génétique quantitative, dendroplasticité et statistique.

2011-2013 : Projet du Ministère de l'écologie, de l'énergie, du développement durable et de la mer, GICC « Mesure du potentiel d'adaptation des arbres forestiers au changement climatique : approches *in-situ* et *ex-situ* sur gradients altitudinaux à l'aide de dispositifs de transplantation croisée » (**GRAAL**) (coordonnateur Philippe Rozenberg) ; activités en génétique quantitative, dendroplasticité et statistique.

5. Collaborateurs

Une partie non négligeable des travaux de recherches est faite au sein de l'unité de recherches « Amélioration, Génétique et Physiologie Forestières » de l'INRA d'Orléans, avec le concours de plusieurs personnes faisant partie de l'**équipe génétique** (coordinatrice C. Bastien). Parmi ses personnes, je soulignerais les suivantes (avec quelques mots clés qui reflètent nos intérêts en commun) : **Philippe Rozenberg** (dendroplasticité, adaptation au changement climatique) ; **Catherine Bastien** (amélioration, résistances aux maladies) ; **Hélène Muranty** (modélisation, SAM) ; **Véronique Jorge** (biologie moléculaire, SAM) ; **Luc Pâques** (amélioration, dendroplasticité) ; **Jean-Charles Bastien** (programme Douglas).

France - UMR BIOGECO, INRA Bordeaux (**A. Raffin, S. Gerber, L. Bouffier**) : projets NOVELTREE et PAF-Dendroplasticité, SAM, dendroplasticité et héritabilité *in situ*.

France - UR EFM, INRA Avignon (**S. Oddou, F. Lefèvre**) : projets PAF-Dendroplasticité et GRAAL, modélisation et héritabilité *in situ*.

France - Unité « Diversité génétique et amélioration des espèces forestières », CIRAD Montpellier (**D. Cros, J.-M. Bouvet**) : héritabilité *in situ*, SAM.

France - UR GAPF, INRA Lusignan (**I. Litrico**) : modélisation de la plasticité phénotypique et participation comme expert dans un projet sur le ray-grass.

Argentine - INTA Bariloche (**A. Martinez-Meier, L. Gallo, M. Pastorino, G. dalla Salda**) : projets ALFA II GEMA et ECOS-SUD Argentine, mise en place de gradients altitudinaux et études de dendroplasticité.

Argentine - Universidad de Buenos Aires (**J.C. Vilardi, B. Saidman**) : projet ALFA II GEMA, travaux de simulation sur l'estimation des paramètres génétiques sans connaissance a priori des relations d'apparement, héritabilité *in situ*.

Australie - Groupe de Génétique Quantitative et Amélioration, CSIRO Canberra (**H. Wu, M. Ivkovich, B. Baltunis**) : modélisation en génétique quantitative et de populations.

Canada - British Columbia Ministry of Forestry Research Branch (**A. Yanchuk, J. Russell, J. King, M. Stoehr**) : stratégies de sélection.

Écosse - Division de Génétique et Biométrie, Institute Roslin de l'Université d'Édimbourg (**J. A. Woolliams**) : projet NOVELTREE, sélection pangénomique.

Espagne - CIS Madeira (**E. Merlo**) : projets CIS Madeira et PICASSO.

Espagne - Département de Biochimie, Génétique et Immunologie de l'Université de Vigo (**A. Caballero**) : projet Université de Vigo, SAM et balayage sélectif.

Portugal – RAIZ, Direcção de Investigação Florestal (**C. Marques**) et Unidade Departamental de Silvicultura e Recursos Naturais, Escola Superior Agrária (**M. M. Ribeiro**) : utilisation de marqueurs pour le suivi de la diversité en sélection.

Suède - Umeå Plant Science Centre (**H. Wu**) : modélisation en génétique quantitative et de populations.

Encadrements

- Co-encadrement MASTER 2 Écosystèmes Terrestres de l'Université d'Orléans de l'ingénieur Guillermina dalla Salda de l'INTA Bariloche (Argentine), de juin 2005 à décembre 2006, soutenance décembre 2006.
- Co-encadrement DIRS de l'Université de Tours de Muriel Thomasset, de septembre 2006 à septembre 2007, soutenance octobre 2007.
- MASTER 2 SRO Statistiques et Recherche Opérationnelle de l'Université d'Orléans de Vincent Garnier, de avril 2010 à septembre 2010, soutenance septembre 2010.
- Co-encadrement MASTER 2 SRO Statistiques et Recherche Opérationnelle de l'Université d'Orléans de Cécile Sauder, de avril 2010 à septembre 2010, soutenance Septembre 2010.
- MASTER 2 Écosystèmes Terrestres de l'Université d'Orléans de Charlène Branger, de septembre 2010 à juin 2011, soutenance prévue juin 2011.

- Thèse d'Alejandro Martinez-Meier (thèse Agro Paris-Tech ENGREF), de Juin 2005 à Janvier 2009. Médaille d'argent 2009 de l'Académie d'agriculture de France.
- Thèse d'Anne-Sophie Sergent (thèse Agro Paris-Tech ENGREF), septembre 2007 et fin prévue pour décembre 2010.
- Thèse d'Manuela Ruiz (thèse Agro Paris-Tech ENGREF), mars 2010 et fin prévue pour décembre 2013.

- Co-encadrement stage postdoctoral de Nahum Sanchez-Vargas, enseignant à l'Université de Michoacan (Mexique), de septembre 2005 à août 2006.
- Stage postdoctoral de Lynda Dib, enseignant ingénieur à l'Université Badji Mokhtar d'Annaba (Algérie), de février 2009 à janvier 2011.
- Stage de Maria M. Ribeiro, enseignant chercheur à l'Escola Superior Agrária de Castelo Branco (Portugal), août 2004.
- Stage de Juan Cesar Vilardi, professeur de l'Université de Buenos Aires (Argentine), de novembre 2004 à avril 2005, et de décembre 2006 à janvier 2007.
- Stage du chercheur Harry Wu, directeur du Groupe de Génétique Quantitative et Amélioration au CSIRO Canberra (Australie), de mai 2008 à octobre 2008.
- Stage du chercheur Milosh Ivkovich, chercheur au CSIRO Canberra (Australie), juillet 2010.

6. Références bibliographiques

- [1] S. E. Aggrey, C. Y. Lin, and K. M. Cheng. Size of breeding populations required for selection programs. *Theoretical and Applied Genetics*, 91(4):553–556, 1995.
- [2] E. W. Andersson, L. Sanchez Rodriguez, and B. Andersson. Group coancestry-controlled selection in a pinus sylvestris l. breeding program. *Theoretical and Applied Genetics*, 99(1/2), 1999.
- [3] A. L. Barabasi and Z. N. Oltvai. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–U15, 2004.
- [4] G. M. Belonsky and B. W. Kennedy. Selection on individual phenotype and best linear unbiased predictor of breeding value in a closed swine herd. *Journal of Animal Science*, 66(5):1124–1131, 1988.
- [5] C. Bessega, B. O. Saidman, M. R. Darquier, M. Ewens, L. Sanchez, P. Rozenberg, and J. C. Vilardi. Consistency between marker- and genealogy-based heritability estimates in an experimental stand of prosopis alba (leguminosae). *American Journal of Botany*, 96(2):458–465, 2009.
- [6] P. Bijma, J. A. M. van Arendonk, and J. A. Woolliams. A general procedure for predicting rates of inbreeding in populations undergoing mass selection. *Genetics*, 154(4):1865–1877, 2000.
- [7] P. Bijma, J. A. M. van Arendonk, and J. A. Woolliams. Predicting rates of inbreeding for livestock improvement schemes. *Journal of Animal Science*, 79(4):840–853, 2001.
- [8] P. Bijma and J. A. Woolliams. Prediction of rates of inbreeding in populations selected on best linear unbiased prediction of breeding value. *Genetics*, 156(1):361–373, 2000.
- [9] J. R. Brisbane and J. P. Gibson. Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. *Theoretical and Applied Genetics*, 91(3):421–431, 1995.
- [10] G. R. Brown, G. P. Gill, R. J. Kuntz, C. H. Langley, and D. B. Neale. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15255–15260, 2004.
- [11] M. G. Bulmer. *The mathematical theory of quantitative genetics*. Clarendon Press.; Oxford; UK, 1980.
- [12] Molly K. Burke, Joseph P. Dunham, Parvin Shahrestani, Kevin R. Thornton, Michael R. Rose, and Anthony D. Long. Genome-wide analysis of a long-term evolution experiment with drosophila. *Nature*, advance online publication:–, September 2010.
- [13] A. Caballero. Developments in the prediction of effective population size. *Heredity*, 73(6):657–679, 1994.
- [14] A. Caballero and M. A. Toro. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research*, 75(3):331–343, 2000.
- [15] M. A. Cleveland, H. D. Blackburn, R. M. Enns, and D. J. Garrick. Changes in inbreeding of us herefords during the twentieth century. *Journal of Animal Science*, 83(5):992–1001, May 2005.
- [16] H. Cochard, T. Holtta, S. Herbette, S. Delzon, and M. Mencuccini. New insights into the mechanisms of water-stress-induced cavitation in conifers. *Plant Physiology*, 151(2):949–954, 2009.

- [17] J. K. Conner. Genetic mechanisms of floral trait correlations in a natural population. *Nature*, 420(6914):407–410, 2002.
- [18] C. R. Darwin. *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray, 6th ed edition, 1876.
- [19] A. Dejardin, J. C. Leple, M. C. Lesage-Descauses, G. Costa, and G. Pilate. Expressed sequence tags from poplar wood tissues - a comparative analysis from multiple libraries. *Plant Biology*, 6(1):55–64, 2004.
- [20] Scott J. Dixon, Michael Costanzo, Anastasia Baryshnikova, Brenda Andrews, and Charles Boone. Systematic mapping of genetic interaction networks. *Annual Review of Genetics*, 43(1):601–625, December 2009.
- [21] T. F. C. Falconer, D. S.; Mackay. *Introduction to quantitative genetics*. Longman Group Limited; Harlow; UK, ed. 4 edition, 1996.
- [22] Ronald Aylmer Fisher. *The genetical theory of natural selection*. Clarendon Press.; Oxford; UK, 1930.
- [23] M. E. Goddard. Optimal effective population size for the global population of black and white dairy cattle. *Journal of Dairy Science*, 75(10):2902–2911, 1992.
- [24] A. T. Groover. What genes make a tree a tree? *Trends in Plant Science*, 10(5):210–214, 2005.
- [25] V. P. Gutschick and H. BassiriRad. Extreme events as shaping physiology, ecology, and evolution of plants: toward a unified definition and evaluation of their consequences. *New Phytologist*, 160(1):21–42, 2003.
- [26] J. Hallander and P. Waldmann. Optimum contribution selection in large general tree breeding populations with an application to scots pine. *TAG Theoretical and Applied Genetics*, 118(6):1133–1142, 2009.
- [27] G. Hammer, M. Cooper, F. Tardieu, S. Welch, B. Walsh, F. van Eeuwijk, S. Chapman, and D. Podlich. Models for navigating biological complexity in breeding improved crop plants. *Trends in Plant Science*, 11(12):587–593, 2006.
- [28] A. Hampe and R. J. Petit. Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, 8(5):461–467, 2005.
- [29] J. L. Hamrick, M. J. W. Godt, and S. L. Sherman-Broyles. Factors influencing levels of genetic diversity in woody plant species. *New Forests*, 6(1-4):95–124, 1992.
- [30] W. G. Hill. Variation in response to selection. *Proceedings of the International Conference on Quantitative Genetics, August 16-21, 1976. V. Theory of response and limits to selection.*, pages 343–365, 1977.
- [31] Deng HongWen, V. Haynatzka, K. Spitze, and G. Haynatzki. The determination of genetic covariances and prediction of evolutionary trajectories based on a genetic correlation matrix. *Evolution*, 53(5):1592–1599, 1999.
- [32] R. Lande. The genetic correlation between characters maintained by selection, linkage and inbreeding. *Genetical Research*, 44(3):309–320, 1984.
- [33] J. C. Leple, R. Dauwe, K. Morreel, V. Storme, C. Lapierre, B. Pollet, A. Naumann, Kang KyuYoung, H. Kim, K. Ruel, A. Lefebvre, J. P. Joseleau, J. Grima-Pettenati, R. de Rycke, S. Andersson-Gunneras, A. Erban, I. Fehrle, M. Petit-Conil, J. Kopka, A. Polle, E. Messens, B. Sundberg, S. D. Mansfield, J. Ralph, and G. (et al) Pilate. Downregulation of cinnamoyl-coenzyme a reductase in poplar: multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell*, 19(11):3669–3691, 2007.

- [34] D. A. Levin and A. C. Wilson. Rates of evolution in seed plants: net increase in diversity of chromosome numbers and species numbers through time. *Proceedings of the National Academy of Sciences of the United States of America*, 73(6):2086–2090, 1976.
- [35] Yang Li, Olga Alda Álvarez, Evert W. Gutteling, Marcel Tijsterman, Jingyuan Fu, Joost A. G. Riksen, Esther Hazendonk, Pjotr Prins, Ronald H. A. Plasterk, Ritsert C. Jansen, Rainer Breitling, and Jan E. Kammenga. Mapping determinants of gene expression plasticity by genetical genomics in *c. elegans*. *PLoS Genetics*, 2(12):e222, 2006.
- [36] A. Martinez-Meier, L. Sanchez, G. Dalla-Salda, L. Gallo, M. Pastorino, and P. Rozenberg. Ring density record of phenotypic plasticity and adaptation to drought in douglas-fir. *Forest Ecology and Management*, 258(5):860–867, 2009.
- [37] A. Martinez-Meier, L. Sanchez, M. Pastorino, L. Gallo, and P. Rozenberg. What is hot in tree rings? the wood density of surviving douglas-firs to the 2003 drought and heat wave. *Forest Ecology and Management*, 256(4):837–843, 2008.
- [38] Martinez Meier, Alejandro G., Sanchez, Leopoldo, Dalla Salda, Guillermina, Pastorino, Mario J.M., Gautry, Jean-Yves, Gallo, Leonardo A., and Rozenberg, Philippe. Genetic control of the tree-ring response of douglas-fir (*pseudotsuga menziesii* (mirb.) franco) to the 2003 drought and heat-wave in france. *Ann. For. Sci.*, 65(1):102, 2008.
- [39] J. M. McNamara and A. I. Houston. State-dependent life histories. *Nature*, 380(6571):215–221, 1996.
- [40] J. Merila and B. C. Sheldon. Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity*, 83(2):103–109, 1999.
- [41] T. H. E. Meuwissen. Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75(4):934–940, 1997.
- [42] Wei Ming, A. Caballero, and W. G. Hill. Selection response in finite populations. *Genetics*, 144(4):1961–1974, 1996.
- [43] Michael M. Miyamoto and M Goodman. Dna systematics and evolution of primates. *Annual Review of Ecology and Systematics*, 21(1):197–220, 1990.
- [44] D. B. Neale and O. Savolainen. Association genetics of complex traits in conifers. *Trends in Plant Science*, 9(7):325, 2004.
- [45] R. J. Petit and A. Hampe. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, 37:187–214, 2006.
- [46] M. Quinton and C. Smith. Comparison of evaluation-selection systems for maximizing genetic response at the same level of inbreeding. *Journal of Animal Science*, 73(8):2208–2212, 1995.
- [47] K. Ritland and S. Travis. Inferences involving individual coefficients of relatedness and inbreeding in natural populations of abies. *Forest Ecology and Management*, 197(1/3):171–180, 2004.
- [48] L. Sanchez, P. Bijma, and J. A. Woolliams. Minimizing inbreeding by managing genetic contributions across generations. *Genetics*, 164(4), 2003.
- [49] L. Sanchez, A. Caballero, and E. Santiago. Maximising polygenic response to selection with an optimal frequency path for a selected major gene. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006*, pages 30–17, 2006.
- [50] L. Sanchez, A. Caballero, and E. Santiago. Palliating the impact of fixation of a major gene on the genetic variation of artificially selected polygenes. *Genetical Research*, 88(2):105, 2006.

- [51] L. Sanchez, C. Garcia, J. Lomas, A. Moreno, B. Nieto, J. Piqueras, C. Salgado, and M. Toro. Lab experiments on optimum contribution selection. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, 2002. Session 19*, pages 1–5, 2002.
- [52] L. Sanchez, A. Yanchuk, and J. King. Gametic models for multitrait selection schemes to study variance of response and drift under adverse genetic correlations. *Tree Genetics & Genomes*, 2007.
- [53] Leopoldo Sanchez. Balanced vs. slightly unbalanced selection. *Heredity*, 84(6):685–691, 2000.
- [54] Leopoldo Sanchez and John A. Woolliams. Impact of nonrandom mating on genetic variance and gene flow in populations with mass selection. *Genetics*, 166(1):527–535, 2004.
- [55] N. M. Sanchez-Vargas, L. Sanchez, and P. Rozenberg. Plastic and adaptive response to weather events: a pilot study in a maritime pine tree ring. *Canadian Journal of Forest Research*, 37(11):2090–2095, 2007.
- [56] E. Santiago and A. Caballero. Effective size of populations under selection. *Genetics*, 139(2):1013–1030, 1995.
- [57] O. Savolainen and T. Pyhajarvi. Genomic diversity in forest trees. *Current Opinion in Plant Biology*, 10(2):162–167, 2007.
- [58] N. Scarcelli, J. M. Cheverud, B. A. Schaal, and P. X. Kover. Antagonistic pleiotropic effects reduce the potential adaptive value of the frigida locus. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43):16986–16991, 2007.
- [59] Carl D. Schlichting, Massimo Pigliucci, C. D. Schlichting, and M. Pigliucci. Phenotypic evolution: A reaction norm perspective. *Phenotypic evolution: A reaction norm perspective*, pages xii+387p, 1998.
- [60] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics*, 31(1):64–68, 2002.
- [61] E. C. Snell-Rood, J. D. Van Dyken, T. Cruickshank, M. J. Wade, and A. P. Moczek. Toward a population genetic framework of developmental evolution: the costs, limits, and consequences of phenotypic plasticity. *Bioessays*, 32(1):71–81, 2010.
- [62] A. K. Sonesson and T. H. E. Meuwissen. Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution*, 32(3):231–248, 2000.
- [63] A. K. Sonesson and T. H. E. Meuwissen. Non-random mating for selection with restricted rates of inbreeding and overlapping generations. *Genetics Selection Evolution*, 34(1):23–39, 2002.
- [64] A. C. Sorensen, P. Berg, and J. A. Woolliams. The advantage of factorial mating under selection is uncovered by deterministically predicted rates of inbreeding. *Genetics Selection Evolution*, 37(1):57, 2005.
- [65] M. Stoehr, A. Yanchuk, Xie ChangYi, and L. Sanchez. Gain and diversity in advanced generation coastal douglas-fir selections for seed production populations. *Tree Genetics and Genomes*, 4(2):193–200, 2008.
- [66] S. J. Tonsor, C. Alonso-Blanco, and M. Koornneef. Gene function beyond the single trait: natural variation, gene effects, and evolutionary ecology in arabidopsis thaliana. *Plant, Cell and Environment*, 28(1):2–20, 2005.
- [67] M. Toro and M. Perez-Enciso. Optimization of selection response under restricted inbreeding. *Genetics, Selection, Evolution*, 22(1):93–107, 1990.

- [68] M. Toro, L. Silio, and M. Perez-Enciso. A note on the use of mate selection in closed moet breeding schemes. *Animal Production*, 53(3):403–406, 1991.
- [69] M. A. Toro and B. M. Nieto. A simple method for increasing the response to artificial selection. *Genetical Research*, 44(3):347–349, 1984.
- [70] G. A. Tuskan, S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R. R. Bhalerao, R. P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G. L. Chen, and D. (et al) Cooper. The genome of black cottonwood, *populus trichocarpa* (torr. & gray). *Science (Washington)*, 313(5793):1596–1604, 2006.
- [71] A. L. Tyler, F. W. Asselbergs, S. M. Williams, and J. H. Moore. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays*, 31(2):220–227, 2009.
- [72] B. Villanueva, J. C. M. Dekkers, J. A. Woolliams, and P. Settar. Maximizing genetic gain over multiple generations with quantitative trait locus selection and control of inbreeding. *Journal of Animal Science*, 82(5):1305–1314, 2004.
- [73] R. P. Wei. Loss of genetic diversity following selection from populations with a family structure. *Silvae Genetica*, 45(2/3):153–159, 1996.
- [74] K. A. Weigel. Controlling inbreeding in modern breeding programs. *Journal of Dairy Science*, 84(Elect. Supplement):E177–E184, 2001.
- [75] C. G. Williams and O. Savolainen. Inbreeding depression in conifers: implications for breeding strategy. *Forest Science*, 42(1):102–117, 1996.
- [76] J. A. Woolliams. Designs and evaluations for managing genetic diversity in breeding programs. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006*, pages 30–01, 2006.
- [77] N. R. Wray and M. E. Goddard. Increasing long-term response to selection. *Genetics, Selection, Evolution*, 26(5):431–451, 1994.
- [78] S. Wright. Evolution and the genetics of populations. volume 3. experimental results and evolutionary deductions. *Evolution and the genetics of populations. Volume 3. Experimental results and evolutionary deductions.*, page 613pp., 1977.
- [79] R. L. Wu and M. Lin. Opinion - functional mapping - how to map and study the genetic architecture of dynamic complex traits. *Nature Reviews Genetics*, 7(3):229–237, March 2006.
- [80] J. B. Zobel, B. J.; Jett. *Genetics of wood production*. Springer-Verlag; Berlin; Germany, College of Forest Resources, North Carolina State University, PO Box 8002, Raleigh, NC 27695-8002, USA., 1995.

Annexe : Curriculum Vitae

Leopoldo SANCHEZ RODRIGUEZ, Research Scientist
INRA Orléans, Unité d'Amélioration, Génétique et Physiologie Forestières (UAGPF)
BP 20619 Ardon 45166 Olivet Cedex, France
Tel: +33 (0)2 38 41 78 14; Fax: +33 (0)2 38 41 78 79
Email: leopoldo.sanchez@orleans.inra.fr

Main research lines: *quantitative genetics methodologies applied to natural and ex situ populations, quantitative genetics of phenotypic plasticity and optimization of breeding strategies for genetic diversity management in forest trees*

Education and employment History

- Since 02/2002 Permanent research scientist at UAGPF, INRA – Orléans, France.
- 02/2001 *Marie Curie Fellow* Researcher (European Commission) at the Department of Biochemistry, Genetics and Immunology, Vigo University, Spain.
- 02/2000 *Marie Curie Fellow* Researcher (European Commission) at the Roslin Institute of Edinburgh, Genetics and Biometry Division, Edinburgh, Scotland.
- 11/1997 *Marie Curie Fellow* Researcher (European Commission) at the Department of Genetics and Tree Physiology, Swedish University of Agricultural Sciences. Umeå, Sweden.
- 05/1997 PhD Thesis in Quantitative Genetics, University of Santiago de Compostela, Spain.

Some relevant research projects

- 2007 *Large collaborative project of the EU 7th Framework Programme: NOVELTREE; activity leader in “Providing strategies to overcome breeding constraints in forest trees”.*
- 2007 *Project of the French National Research Funding Agency: DRYADE; researcher on genetics of plasticity.*
- 2007 *Project of the French Agricultural Funding Agency: Dendroplasticité; researcher on quantitative genetics of phenotypic plasticity.*
- 2007 *Large collaborative network of the 6th Framework Programme: TREEBREDEX; activity coordinator in long-term breeding and statistics workshops.*
- 2006 *Europe-Latin America research project Alfa: GEMA Genetics of Wood Formation; co-supervisor of 5 postgraduate and postdoctoral researchers.*
- 2005 *Projet Innovant du Département INRA EFPA: “Neutral markers and relatedness in the estimation of adaptive genetic variability in natural forest”; coordinator.*
- 2004 *Projet Innovant du Département INRA EFPA: “Modélisation: un outil d’aide au choix des stratégies optimales, pour la gestion des populations d’amélioration, et la constitution de réseaux de conservation des ressources génétiques”; coordinator*

Main expertise

- 07/2009 Award McMaster Fellowship (Australia) to expertise on allele-based simulation models, CSIRO, Canberra, Australia;
- 04/2008 Expertise in quantitative genetics simulation models, CSIRO, Canberra, Australia;
- 06/2005 & 08/2003 Expertise in quantitative genetics analysis and simulation models, British Columbia Ministry of Forestry, Victoria, Canada;
- 04/2003 Expertise in quantitative genetics simulation models, Roslin Institute of Edinburgh, Scotland.

Some relevant publications since research scientist at INRA

Martinez Meier A, **Sánchez L**, Dalla Salda G, Gallo L, Pastorino M & Rozenberg P. 2009. Ring density record of phenotypic plasticity and adaptation to drought in Douglas-fir. ***Forest Ecology and Management*** 258: 860–867.

- Bessegga C F, Saidman B O, Darquier M R, Ewens M, **Sánchez L**, Rozenberg P & Vilardi J C. 2009. Consistency between marker- and genealogy-based heritability estimates in an experimental stand of *Prosopis alba* (Leguminosae). *American Journal of Botany*, 96:458-465.
- Martínez Meier A, **Sánchez L**, Pastorino M, Gallo L & Rozenberg P. 2008. What is hot in tree rings? The wood density of surviving Douglas-firs to the 2003 drought and heat wave. *Forest Ecology and Management*, 256: 837–843.
- Martínez Meier A, **Sánchez L**, Dalla Salda G, Pastorino M, Gautry JY, Gallo L & Rozenberg P. 2008. Genetic control of the tree-ring response of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) to the 2003 drought and heat-wave in France. *Annals of Forest Science*, 65: 102-110.
- Sánchez-Vargas N, **Sánchez L** & Rozenberg P. 2007. Plastic and adaptive response to weather events: a pilot study in a maritime pine tree-ring. *Canadian Journal of Forest Research*, 37(11): 2090-2095.
- Sánchez L**, Yanchuk, AA. & King JN. 2007. Gametic models for multitrait selection schemes to study variance of response and drift under adverse correlations. *Tree Genetics & Genomes* (DOI 10.1007/s11295-007-0101-5).
- Stoehr M, Yanchuk AA, Xie CY & **Sánchez L**. 2007. Gain and diversity in advanced generation coastal Douglas fir selections for seed productions populations: to subline or not to subline. *Tree Genetics & Genomes* (DOI 10.1007/s11295-007-0100-6).
- Sánchez L**, Santiago E & Caballero A. 2006. Palliating the impact of fixation of a major gene on the genetic variation of artificially selected polygenes. *Genetical Research*, 87: 1-14.
- Wenes AL, Bourguet D, Andow DA, Courtin C, Carré G, Lorme P, **Sánchez L** & Augustin S. 2006. Frequency and fitness cost of resistance to *Bacillus thuringiensis* in *Chrysomela tremulae* (Coleoptera: Chrysomelidae). *Heredity*, 97: 127-134.
- Sánchez L**, Santiago E & Caballero A. 2006. Maximising polygenic response to selection with an optimal frequency path for a selected major gene. *Proc. 8th World Cong. Genet. Appl. Livest. Prod.*, 30-17: 1-4.
- Sánchez L** & Wolliams JA. 2004. Impact of non-random mating on genetic variance and gene flow in populations with mass selection. *Genetics*, 166: 527-535.
- Sánchez L**, Bijma P & Wolliams JA. 2003. Minimizing inbreeding by managing genetic contributions across generations. *Genetics*, 164: 1589-1595.
- Ribeiro MM, **Sánchez L**, Ribeiro C, Carocha V, Araújo JA, Borralho N & Marques CM. A case study of *Eucalyptus globulus* fingerprinting for breeding. Submitted to *Annals of Forest Science*.
- Yanchuk A & **Sánchez L**. Multivariate selection under adverse genetic correlations: impacts on population sizes and selection strategies in forest tree breeding. Submitted to *Tree Genetics & Genomes*.
- Brachet S, Mariette S & **Sánchez L**. Sampling genetic diversity in structured natural populations. Submitted to *Conservation Genetics*.

R É P U B L I Q U E F R A N Ç A I S E

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
UNIVERSITÉ D'ORLÉANS

HABILITATION A DIRIGER DES RECHERCHES

Vu le code de l'éducation ;

Vu le décret n° 84-573 du 5 juillet 1984 modifié relatif aux diplômes nationaux de l'enseignement supérieur ;

Vu le décret n° 2002-481 du 8 avril 2002 relatif aux grades et titres universitaires et aux diplômes nationaux ;

Vu l'arrêté du 23 novembre 1988 modifié relatif à l'habilitation à diriger des recherches, notamment son article 5 ;

Vu l'arrêté du 21 décembre 1989 modifié fixant la liste des établissements d'enseignement supérieur autorisés à délivrer, seuls, l'habilitation à diriger des recherches ;

Vu le procès-verbal du jury attestant que l'intéressé a présenté ses travaux, le 25 novembre 2010, devant un jury présidé par ERIC LAINE, Professeur des Universités et composé de JEAN-MARC BOUVET, Chercheur, GILLES CHARMET, Chargé de recherche, MARTINE DECOVILLE, Maître de Conférences, CHRISTINE DILLMANN, Professeur des Universités, ALAIN FRANÇ, Directeur de Recherche, FRANÇOIS LEFEVRE, Directeur de Recherche ;

Vu la délibération du jury ;

Le diplôme d' **HABILITATION À DIRIGER DES RECHERCHES SCIENCES DE LA VIE**

est délivré à **M. LEOPOLDO SANCHEZ-RODRIGUEZ** né le 2 novembre 1966 à ST JACQUES DE COMPOSTELLE (ESPAGNE)
pour en jouir avec les droits et prérogatives qui y sont attachés.

Fait à Orléans, le 31 janvier 2011

Le titulaire

Le Président

Youssef TOURE

Le Recteur d'Académie,
Chancelier des universités

Paul CANTONI

N°
ORLÉANS 8269340
/2011200901838

Ce document est imprimé sur du papier filigrané avec des encres réactives et ne peut être ni gommé ni raturé.

Il est affecté d'un numéro (N° xxx xxx./....) dont vous pouvez vérifier l'exactitude auprès du **Service Diffusion de l'Imprimerie Nationale** :

☎ 03 27 93 70 84 ou 03 27 93 70 97

00 (33) 3 27 93 70 84 ou 3 27 93 70 97 (de l'étranger)