



HAL
open science

Effets de la sélection sur la diversité et la différenciation génétique moléculaire : résultats de simulations et application au pin maritime pour les gènes de la lignification

Pauline Garnier-géré, Frédéric Austerlitz, Frank Bedon, Antoine Kremer,
Christophe Plomion

► To cite this version:

Pauline Garnier-géré, Frédéric Austerlitz, Frank Bedon, Antoine Kremer, Christophe Plomion. Effets de la sélection sur la diversité et la différenciation génétique moléculaire : résultats de simulations et application au pin maritime pour les gènes de la lignification. 5. Colloque national du BRG, Nov 2004, Lyon, France. hal-02762648

HAL Id: hal-02762648

<https://hal.inrae.fr/hal-02762648v1>

Submitted on 4 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.

Effets de la sélection sur la diversité et la différenciation génétique moléculaire : résultats de simulations et application au pin maritime pour les gènes de la lignification

Pauline GARNIER-GÉRÉ^{(1)*}, Frédéric AUSTERLITZ⁽²⁾,
Frank BEDON⁽³⁾, Antoine KREMER⁽¹⁾, Christophe PLOMION⁽¹⁾

⁽¹⁾ INRA, UMR 1202, Équipe de génétique, 69 route d'Arcachon,
33612 Cestas Cedex, France

⁽²⁾ Laboratoire Écologie, systématique et évolution, Université Paris-Sud,
UPRESA-CNRS 8079, Bat. 362, 91405, Orsay Cedex, France

⁽³⁾ Adresse actuelle : Centre de recherche en biologie forestière, Université
Laval, Pavillon Marchand, Laboratoire 2152, Québec, Canada G1K 7P4

Abstract: Effects of selection on molecular diversity and differentiation patterns of genes involved in the expression of a quantitative trait: simulation results and case study of the lignin biosynthesis genes in maritime pine (*Pinus pinaster* Ait.). Molecular diversity patterns for genes involved in adaptive trait variation have been investigated in a forest tree species following two different approaches. In the experimental approach, which targeted lignin biosynthesis and other wood formation candidate genes, in a sample of *P. pinaster* populations, a large variation of nucleotide diversity estimates was found among gene fragments (from 0.00193 to 0.00773), with one single nucleotide polymorphism every 40 base pairs on average. Different neutrality tests based either on the comparison of nucleotide diversity estimates, or on site frequency distributions showed significant departures from neutral expectations for most genes studied. The patterns observed within a more extensively sampled population were interpreted as directional selection events that could have occurred either recently or in a much more distant past, depending on the genes. Using all populations, the nucleotide differentiation of 2 out of 5 candidate genes was far greater than the upper bound of differentiation values (F_{st}) obtained for neutral or anonymous markers. This could be interpreted as diversifying selection between locally adapted populations. In the simulation approach, we explore the evolution of molecular differentiation at genes controlling an adaptive trait in a subdivided population, and the genetic hitchhiking effects at surrounding neutral markers, localised at increasing recombination rates.

Correspondance et tirés à part : pauline@pierroton.inra.fr

Previous works have studied the characteristic features of the hitchhiking process, such as a reduction of diversity around the selected locus and an excess of high-frequency derived mutations, which have both been used for testing departures from neutrality. However, the consequences of genetic hitchhiking on differentiation between populations at linked neutral loci have received much less interest. The few existing models show that a substantial increase in differentiation can occur at neutral loci for low levels of migration and for recombination rates smaller than selection coefficients. The problem is that these models are usually restricted to one selected locus and a small number of alleles, or to a small number of populations. Thus, the question of the importance of hitchhiking effects in natural populations remains, especially in the case of selection affecting polygenic traits, and for species with high migration rates. In our model, the trait is submitted to both stabilizing selection within population and diversifying selection between populations. We study the impact of these two selective forces on the effect of genetic hitchhiking on differentiation, focusing on situations that are representative of allogamous tree species with large population sizes and high migration rates. The results of both approaches are discussed in connection with problems of validating the role of candidate genes using association studies in natural or breeding populations, and that of screening genomes for signatures of selection.

nucleotide diversity / selection / hitchhiking / *Pinus pinaster* / lignification

Résumé : Pour mieux comprendre l'organisation de la diversité moléculaire de gènes impliqués dans la variation d'un caractère adaptatif, nous avons suivi deux approches. Une approche expérimentale sur des gènes de la lignification chez le pin maritime nous a permis de mettre en évidence une grande variation de diversité nucléotidique selon les gènes (de 0,00193 à 0,00773), avec un polymorphisme nucléotidique toutes les 40 paires de bases en moyenne. Nous avons observé des écarts significatifs à la neutralité sélective pour la plupart des gènes étudiés, ainsi qu'une différenciation nucléotidique importante pour certains d'entre eux, ce qui a pu être interprété comme la signature d'événements de sélection. En parallèle, nous avons utilisé un modèle simulant l'évolution de la différenciation de gènes à effets quantitatifs (QTL) contrôlant un caractère sous sélection, et celle de marqueurs neutres liés génétiquement à ces QTL. Dans ces simulations, dix QTL indépendants contrôlent un caractère sous sélection dans 25 populations (sélection optimisante intrapopulation, et diversifiante interpopulations). Nous montrons que les effets d'auto-stop génétique affectent la différenciation des marqueurs neutres, en moyenne dans une fenêtre de déséquilibre de liaison de 0,5 cM autour du QTL, si l'une au moins des deux forces sélectives (intra ou interpopulations) est importante. Pour certains locus, ces effets peuvent s'étendre sur une fenêtre dix fois plus grande.

diversité nucléotidique / sélection / auto-stop génétique / *Pinus pinaster* / lignification

1. INTRODUCTION

La diversité intra et interpopulations de gènes impliqués dans des caractères adaptatifs a surtout été étudiée chez l'homme et quelques espèces modèles ou domestiquées (drosophile, arabidopsis, maïs...). Ces gènes conditionnent les changements adaptatifs et devraient être plus souvent la cible de programmes de conservation, alors que classiquement, des marqueurs neutres sont étudiés, qui enregistrent l'histoire démographique (au sens large) des populations [12]. Pour mieux comprendre le niveau et la structure de la diversité nucléotidique de ces gènes candidats, nous présentons deux approches complémentaires, l'une expérimentale et l'autre par simulation.

L'approche expérimentale a pour objectifs de quantifier la diversité nucléotidique de gènes candidats impliqués dans la lignification chez une espèce forestière, de la comparer à celle de marqueurs neutres, et de tester les écarts éventuels à la neutralité de ces gènes grâce aux méthodes d'analyses des patrons de diversité nucléotidique, qui permettent d'aborder l'étude de l'adaptation au niveau moléculaire [8]. La disponibilité croissante de séquences complètes de gènes candidats et de séquences partielles d'ESTs (*Expressed sequenced tags*) a contribué largement à l'application et au développement théorique de ces méthodes [20].

L'approche par simulation vise à tester l'influence de la sélection sur l'évolution et l'organisation de la diversité multilocus intragénome. En particulier, nous avons essayé de quantifier les effets d'entraînement (ou d'auto-stop génétique) de gènes impliqués dans la variation d'un caractère quantitatif (QTLs), lui-même soumis à la sélection. Ces effets peuvent influencer à la fois la diversité et la différenciation de marqueurs neutres qui sont liés physiquement aux QTLs. Parmi les conséquences les mieux étudiées de l'auto-stop génétique sur le plan théorique, on peut citer la réduction de la diversité [27] et la production d'un excès d'haplotypes portant des mutations dérivées [7] autour du locus sélectionné. Par contre, l'effet de l'auto-stop sur la différenciation entre populations a été beaucoup moins étudié. Les quelques modèles développés sont limités à un seul locus sélectionné et un petit nombre d'allèles, ou bien à quelques populations seulement [36], [1]. Donc la question de l'importance réelle de ce phénomène en populations naturelles reste entière, en particulier dans le cas d'un grand nombre de gènes impliqués dans le déterminisme d'un caractère adaptatif soumis à la sélection. La complexité de la modélisation d'un tel problème justifie l'approche par simulations.

2. MATÉRIEL ET MÉTHODES

2.1. Populations et gènes

Dans l'approche expérimentale, neuf populations de pin maritime représentatives de son aire naturelle de distribution ont été choisies parmi les collections disponibles de l'INRA (France métropolitaine (Aquitaine), Corse (Porto-Vecchio, Vivario, Zonza), Portugal (Leria Mata, Leria Velka), Maroc (Puntaceras, Tamjoute), Tunisie (Tabarka)). La région Aquitaine a été échantillonnée de façon plus exhaustive et considérée comme une seule population non structurée, sur la base de résultats antérieurs montrant une absence de différenciation aux niveaux moléculaire et phénotypique [2], [26].

L'ADN de 24 à 75 mégagamétophytes (tissu haploïde) a été amplifié et séquencé pour 14 fragments de 300 à 600 paires de bases, selon le protocole décrit par [23]. Ces fragments proviennent de huit gènes candidats impliqués dans la lignification et la formation du bois (tabl. I). Ces gènes ont été choisis en fonction de la présence de séquences orthologues dans les bases d'ESTs de *P. pinaster* et de *P. taeda* (disponibles dans GENBANK). Les séquences obtenues ont été alignées et la validité des polymorphismes détectés a été vérifiée en visualisant les chromatogrammes correspondants à l'aide du logiciel Sequencher (Version 4.0.5, GeneCodes Corporation).

2.2. Analyses des séquences

Différents estimateurs de diversités nucléotidiques et haplotypiques ($H_d = (n/(n-1))/(1 - \sum p_i^2)$, [40]) ont été obtenus sur l'échantillon total de gamètes grâce au logiciel DNAsp (version 4.0, [34]) : θ [39], basée sur le nombre de sites polymorphes, $\theta\pi$ [29], le nombre moyen de différences entre paires de séquences, et θ_H [7], qui est pondérée par l'homozygotie des mutations dérivées. Sous les hypothèses de neutralité, d'équilibre démographique, de panmixie et de mutation selon un modèle en sites infinis, toutes ces statistiques estiment le taux de mutation de la population, $\theta = 4N_e\mu$, avec N_e la taille efficace et μ le taux de mutation par génération [28]. Les taux de recombinaison intragène ont été estimés par la méthode de Hudson (2001) [17].

Tableau I : Nombre de sites polymorphes identifiés, diversité nucléotidique moyennes par site et diversité haplotypique des fragments de gènes étudiés (d'après [28]).

Nom des enzymes codées par les gènes	Abbrévia-tions	Nombre de fragments	Accessions GENBANK	Nb ⁽⁴⁾	Longueur totale séquencée (pb)	Nombre de sites polymorphes	Nombre de substitutions	Nombre d'Indels	$\pi_m^{(5)} \times 10^3$ (écart-type)	Diversité haplotypique (écart-type)
Phénylalanine ammonia-lyase 1 ⁽¹⁾	<i>PAL-1</i>	2	BX784397 + soumise	23, 69	908	17	10	7	5,301 (3,009)	1,0000 (0,0137)
Cinnamate-4-hydroxylase	<i>C4H</i>	3	BX784398 + soumise	18, 23, 66	1 243	41	25	16	7,727 (4,172)	1,0000 (0,0202)
4-hydroxy-cinnamate CoA ligase 1 ⁽²⁾	<i>4CL-1</i>	1	BX784388	52	549	31	22	9	7,618 (4,269)	0,955 (0,013)
Caféoyl-CoA 3-O-méthyl-transférase	<i>CCoAOMT</i>	1	BX784390	21	509	7	7	0	2,675 (1,920)	0,842 (0,075)
Alcool cinnamylique déshydrogénase	<i>CAD</i>	2	BX784386 + BX784387	67, 70	1 127	18	8	10	2,825 (1,640)	0,870 (0,003)
Laccase 4 ⁽³⁾	<i>LAC-4</i>		Soumise	69	522	24	23	1	7,751 (4,343)	0,344 (0,075)
Laccase 2 ⁽³⁾	<i>LAC-2</i>	3	Soumise	18, 21, 23	1 633	24	5	19	1,930 (1,165)	0,947 (0,044)
Arabino-galactane présumée	<i>AGP</i>	1	BX784280	23	321	6	4	2	6,553 (4,278)	0,7391 (0,1015)
	TOTAL ⁽⁶⁾	13		518	6 794	168	104	64	5,759 (3,336)	0,8748 (0,0360)

(1) D'après l'homologie avec les séquences publiées dans Dvornyk *et al.* 2002 [4] (2) d'après la nomenclature donnée par Zhang et Chiang (1997) [41] ; (3) d'après la nomenclature publiée dans Sato *et al.* (2001) [35]; (4) nombre de séquences pour chaque fragment ; (5) Diversité nucléotidique moyenne par site (écart-type) ; (6) ou moyenne pour les valeurs de diversités.

Nous avons utilisé plusieurs tests de neutralité, basés soit sur l'hétérogénéité des taux d'évolution entre différents locus ou classes de mutations (test HKA [15]), ou sur la comparaison des différents estimateurs de diversité nucléotique (tests de Tajima [37], de Fu [9], et test H de Fay et Wu [7]). Ces derniers sont sensibles à une structuration en populations et n'ont été utilisés que pour la population Aquitaine. Une analyse de variance moléculaire (logiciel Arlequin, version 2.0 [6]) sur les matrices de distances entre haplotypes a permis d'estimer la différenciation entre populations pour les fragments de gène les plus polymorphes (paramètre N_{st} , [28]) et pour chaque site suffisamment polymorphe (fréquence des allèles $> 10\%$) (paramètre F_{st}). Cette analyse a été réalisée pour les populations comprenant au moins 5 séquences avec en moyenne 9 à 10 séquences par population.

2.3. Simulations

Dans l'approche par simulations, le logiciel METAPOPOP [25], [21] a permis de modéliser 25 populations de 500 individus échangeant des migrants par graines selon un modèle en île. Nous avons supposé un fort taux de migration entre populations ($N_e m = 10$, avec m le taux de migration), pour se rapprocher de la situation de nombreuses espèces forestières allogames [14]. Le phénotype des individus pour un caractère adaptatif est contrôlé par dix gènes à effet quantitatif (QTL) indépendants. Chaque QTL est situé sur un chromosome différent, et est entouré de 18 gènes marqueurs neutres (9 de chaque côté du QTL) situés à des distances croissantes de recombinaison selon une progression géométrique (allant de $2,5 \cdot 10^{-7}$ cM à 25 cM). Un chromosome supplémentaire comprenant 19 locus disposés de la même façon, mais sans QTL, a été inclus dans la construction des génotypes, afin de disposer d'une référence neutre.

Les populations ont été soumises à quatre scénarii de sélection phénotypique différents. Ces scénarii combinent une sélection stabilisatrice intrapopulation, forte ou faible, à une sélection diversifiante entre populations, forte ou faible, le long d'un gradient de valeurs phénotypiques d'optimum pour le caractère adaptatif [21]. Le taux de mutation était de 10^{-5} pour chacun des 209 locus (dix QTL + 199 gènes neutres). Nous avons simulé tout d'abord une grande population sans sélection pendant 400 000 générations, pour atteindre un quasi-équilibre mutation/dérive/recombinaison. Cet équilibre a été vérifié en comparant la courbe expérimentale des corrélations entre

gènes (r^2) en fonction du taux de recombinaison, à celle prédite analytiquement pour des allèles dont la fréquence est comprise entre 0,05 et 0,95 [16]. Les individus issus de cette grande population ont ensuite été répartis dans 25 populations auxquelles la sélection est appliquée. L'impact des deux forces de sélection, inter et intrapopulation, sur le niveau de différenciation dans les régions entourant les QTL, comme conséquence de l'auto-stop génétique, a alors été étudié. Dix répétitions de chaque scénario de sélection ont été simulées, pour éviter un biais lié à l'organisation génomique correspondant à la situation de départ. Les valeurs de différenciation ont été moyennées sur toutes les positions le long des chromosomes et l'ensemble des répétitions, au sein d'échantillons de 50 individus tirés au hasard dans chacune des 25 populations.

3. RÉSULTATS ET DISCUSSION

3.1. Diversité nucléotidique des gènes candidats et écarts à la neutralité

Les diversités nucléotidiques moyennes par site (π_m) observées dans les fragments de gènes étudiés sont importantes, en comparaison à d'autres organismes (homme [13]; maïs [38]; soja [42]). Au total, 168 sites polymorphes ont été identifiés, qui étaient soit des mutations ponctuelles (SNPs), soit des insertions/délétions (Indels), ce qui correspond à un site polymorphe toutes les 85 paires de bases en moyenne (tabl. I). On ne peut pas affirmer que ces estimations soient représentatives du génome du pin ; elles reflètent plutôt le choix initial de fragments *a priori* plus polymorphes se situant dans les parties 3' des gènes.

Cependant, une amplitude de variation d'un facteur 10 existe selon les fragments séquencés pour π_m , ainsi que pour la diversité allélique ou haplotypique (tabl. I). Celle-ci reflète la plus ou moins grande importance des déséquilibres de liaison entre sites intragène [11]. En effet, un nombre important d'haplotypes observé dans un échantillon signifie que la recombinaison aura eu le temps de casser les haplotypes ancestraux et donc de créer des haplotypes dérivés. Dans ce cas, les déséquilibres de liaison diminueront rapidement avec la distance le long du chromosome, illustrant ainsi le nombre important de généalogies différentes dans l'échantillon [31]. Par exemple, les 23 substitutions polymorphes du gène LAC-4 sont en déséquilibre de liaison complet, ce qui correspond à une diversité haplotypique trois fois inférieure à celle des autres gènes, et à un

taux de recombinaison proche de zéro (tabl. I et II). À l'inverse, les gènes PAL-1 et 4CL-1 présentent une diversité haplotypique proche de (ou égale à) 1, et sont associés à une diminution très rapide des déséquilibres de liaison entre sites intragène, et à un taux de recombinaison beaucoup plus important (tabl. II et [11]).

Tableau II : Tests de neutralité sélective (* : test significatif à $P < 5\%$) et estimation des taux de recombinaison au sein de la population Aquitaine (d'après [11]).

Gènes ⁽¹⁾	Fragments ⁽²⁾	Nb ⁽³⁾	Nombre de séquences	D de Tajima	F_s de Fu	R_m ⁽⁴⁾	Taux de la recombinaison / gène ⁽⁵⁾ ($R = 4N_e r$)	H de Fay et Wu	test HKA (valeur du X^2) ⁽⁶⁾
PAL-1	PAL-F1	9	33	0,476	- 1,883	1	1,6	0,441	na
PAL-1	PAL-F1,F2	12	12	0,524	- 8,814**	2	8,5	2,212	0,042
C4H	C4H-F1	15	34	- 0,704	- 3,932	2	4,5	- 0,478	na
4CL-1		14	28	- 0,251	- 7,859**	7	7,3	- 1,720	3,201
CAD		10	32	0,465	- 0,717	2	2,0	1,145	6,182*
LAC-4		23	34	- 0,571	5,551**	0	0,0	- 4,877*	0,021

(1) Abréviations suivant le tableau 1 ; (2) quand cela est indiqué, les tests ont été réalisés seulement pour les fragments disponibles ayant un nombre de séquences plus important ; (3) nombre de sites polymorphes ; (4) nombre minimum d'événements de recombinaison d'après le test de Hudson et Kaplan (1985) ; (5) estimé par la méthode de Hudson (2001) [16] ; (6) le test HKA (Hudson, Kreitman, Aguadé 1989, [15]) a été réalisé en comparant parties codante et non codante des fragments séquencés ;* : test significatif à 5 %, ** : test significatif à 1%.

Des écarts significatifs à la neutralité sélective ont été mis en évidence pour 4 des 5 gènes étudiés dans la population Aquitaine (tabl. II). Dans le cas du gène PAL-1, pour le plus long fragment (tabl. II), le test de Fu est hautement significatif, avec une valeur négative de la statistique F_s (- 8,814), ce qui signifie que le nombre d'haplotypes obtenu est supérieur à ce qui est attendu sous l'hypothèse de neutralité sélective et d'équilibre démographique. Inversement, pour le gène LAC-4, ce nombre est significativement inférieur à l'attendu neutre, donnant une valeur de F_s positive (5,551). Ce dernier cas est cohérent avec l'observation d'un faible nombre d'haplotypes, associé à des déséquilibres entre sites importants sur une longue distance, ce qui est une indication d'un effet possible de la sélection sur le gène considéré [18]. De plus, le gène LAC-4 montre un haplotype majoritaire et un excès d'haplotypes rares proches de cet haplotype et portant des mutations dérivées (test de Fay et Wu significatif, tabl. II), ce qui pourrait être le signe d'un événement de sélection positive récent [33].

Déséquilibre de liaison et cartographie fine

La question se pose de savoir si les écarts à la neutralité observés ne peuvent pas être dus à des événements démographiques comme des expansions ou des goulots d'étranglements, l'équilibre démographique étant une autre hypothèse majeure du modèle neutre. Des éléments historiques sur la population d'Aquitaine (nombre de générations depuis la fondation [32], répartition continue, faible altitude, uniformité du milieu...) et l'observation d'une diversité neutre élevée pour des marqueurs AFLP et microsattellites [3], [26], permettent de supposer qu'il n'y a eu ni un effet de fondation fort, ni une réduction importante de la taille de la population au cours de son évolution. D'autre part, les écarts à la neutralité observés sont de sens opposés pour certains gènes, ce qui conforte l'hypothèse d'effets de sélection particuliers à chaque gène, plutôt que la conséquence d'événements démographiques qui auraient affecté le génome de façon plus uniforme.

Concernant la diversité des séquences dans l'ensemble de l'aire de répartition, une forte structuration géographique a été observée pour certains gènes (N_s de 26 % et 42 % pour les gènes CAD et C4H-F1 (tabl. III). La distribution des valeurs de différenciation entre populations aux sites les plus polymorphes (F_{st} intragène) a ensuite été comparée entre les différents gènes. Les gammes de valeurs de F_{st} aux différents sites polymorphes sont très élevées pour certains gènes (31 % à 88 % pour les sites du gène C4H-F1, - 4,7 % à 44,4 % pour le gène CAD, tabl. III), et dépassent celles des bornes supérieures d'un intervalle représentant les valeurs de différenciation estimées pour des marqueurs neutres microsattellites ou bien des marqueurs anonymes AFLP sur les mêmes populations. Ces résultats pourraient indiquer le signe d'une sélection divergente entre les populations au niveau moléculaire, alors que dans le cas du gène C4H, aucun test intrapopulation n'était significatif.

Tableau III : Différenciation nucléotidique entre populations pour les gènes étudiés (d'après [11]).

Gènes ⁽¹⁾	Nombre de séquences	Nombre de populations	Nst ⁽²⁾	Proba ⁽³⁾	Min ⁽⁴⁾	Max ⁽⁵⁾
PAL-F1	67	7	0,13288	0,00196	3,4	24,9
C4H-F1	61	6	0,41709	< 0,00001	31,3	88,3
4CL-1	46	5	0,17597	< 0,00001	- 4,0	53,5
CAD	64	7	0,25778	< 0,00001	- 4,7	44,4
LAC-4	68	7	0,04117	0,16422	4,1	7,0

(1) Abréviations et fragments suivant les tableaux 1 et 2 ; (2) différenciation nucléotidique entre populations ; (3) probabilité qu'une valeur randomisée obtenue par bootstrap (1 000 répétitions) soit supérieure à la valeur observée ; (4) valeur minimale de F_{st} pour les sites polymorphes du fragment (en %) ; (5) valeur maximale de F_{st} pour les sites polymorphes du fragment (en %).

3.2. Importance des effets d'auto-stop génétique sur la différenciation de marqueurs neutres

Dans les simulations, la différenciation (G_{st} , [30]) a été estimée aux différents QTL ainsi qu'aux locus neutres avoisinants sur une période de 15 000 générations. La figure 1 illustre le scénario 1, où une sélection optimisante intrapopulation forte ($w^2 = 1$, avec $1/w^2$ le paramètre représentant l'intensité de la sélection optimisante), est combinée à une sélection diversifiante interpopulation forte (représentée par une variance phénotypique entre valeurs moyennes optimales $V_{Z_{opt}} = 5$).

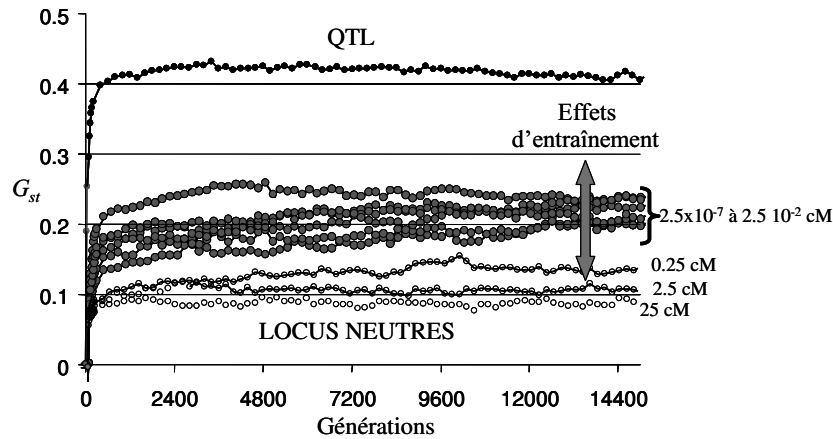


Figure 1 : Évolution des valeurs de G_{st} aux QTL et aux marqueurs neutres situés à des distances croissantes de recombinaison dans un scénario de sélections fortes intra et interpopulation (scénario 1). Chaque courbe est la moyenne de 10 chromosomes sur 10 répétitions.

Quel que soit le scénario considéré, les valeurs moyennes de G_{st} sur les 10 QTL et pour les 10 répétitions atteignent un équilibre rapidement, en quelques centaines de générations, ce qui correspond au temps écoulé depuis la recolonisation consécutive à la dernière période glaciaire pour bon nombre d'espèces d'arbres forestiers. Comparée à l'attendu neutre pour le paramètre G_{st} , qui est d'environ 2,3 % dans un modèle en île (avec $N_e * m = 10$), une différenciation moyenne très forte, de l'ordre de 40 % est observée pour le scénario 1 aux QTL (fig. 1). Ces valeurs sont d'environ 15 % quand la sélection diversifiante est plus faible ($V_{Z_{opt}} = 0,5$, et $1/w^2 = 1$, scénario 2), ce qui est encore largement supérieur à l'attendu neutre, et à 6 % pour le scénario 3 (sélection intrapopulation faible avec $1/w^2 = 0,02$ et sélection interpopulation forte

Déséquilibre de liaison et cartographie fine

avec $V_{z_{opt}} = 5$). Ces résultats sont cohérents avec ceux obtenus par Le Corre et Kremer [21] pour des paramètres d'entrée du modèle similaires. Il faut noter que l'on s'est placé ici dans le cadre d'un modèle *a priori* plus général que celui qui correspondrait au cas du pin maritime, avec des flux de gènes plus importants entre populations, conduisant à une différenciation neutre de l'ordre de 2,5 %. Cependant, la faible migration efficace du scénario 1 conduit à une différenciation aux locus qui ne sont pas entraînés de l'ordre de 10 %, ce qui est cohérent avec les valeurs observées chez le pin maritime pour la diversité neutre [26].

L'observation des valeurs moyennes de G_{st} aux locus neutres situés à des distances croissantes de recombinaison par rapport aux QTL permet de quantifier jusqu'à quelle distance des effets d'entraînement sur la différenciation sont détectables le long du chromosome. De tels effets sont observés en moyenne dans une fenêtre de 0,5 cM autour des QTL, pour les scénarios 1 à 3, caractérisés par une sélection optimisante forte intrapopulation et une sélection diversificatrice forte ou faible interpopulations (fig. 1 et résultats non montrés). Les effets s'installent aussi très rapidement en moyenne, en quelques centaines de générations, suivant ainsi l'évolution du G_{st} aux QTL.

Cependant, on observe également une forte hétérogénéité aux différents QTL et aux locus neutres avoisinants, tant pour l'évolution de la différenciation, que pour l'intensité des effets d'entraînements (fig. 2-A, B, C), qui s'explique en partie par la distribution des déséquilibres de liaison entre les différents gènes le long des chromosomes. Dans le cas où les effets d'entraînement apparaissent, on les observe parfois dans une fenêtre 10 fois supérieure à la fenêtre moyenne autour du QTL, jusqu'à 5 cM. La figure 2-A montre aussi que sur une période de 150 à 200 générations, représentative de la situation actuelle des espèces forestières depuis la recolonisation postérieure aux dernières glaciations, on n'atteint pas l'équilibre pour de nombreux locus, contrairement aux patrons moyens de la figure 1. Par ailleurs, on peut noter des situations où il n'y a pas d'effets d'entraînement, malgré une sélection forte agissant sur le caractère. Ceci pourrait s'expliquer par le fait que les QTL sont en équilibre de liaison avec les locus neutres liés au moment où la sélection est appliquée dans la métapopulation [10].

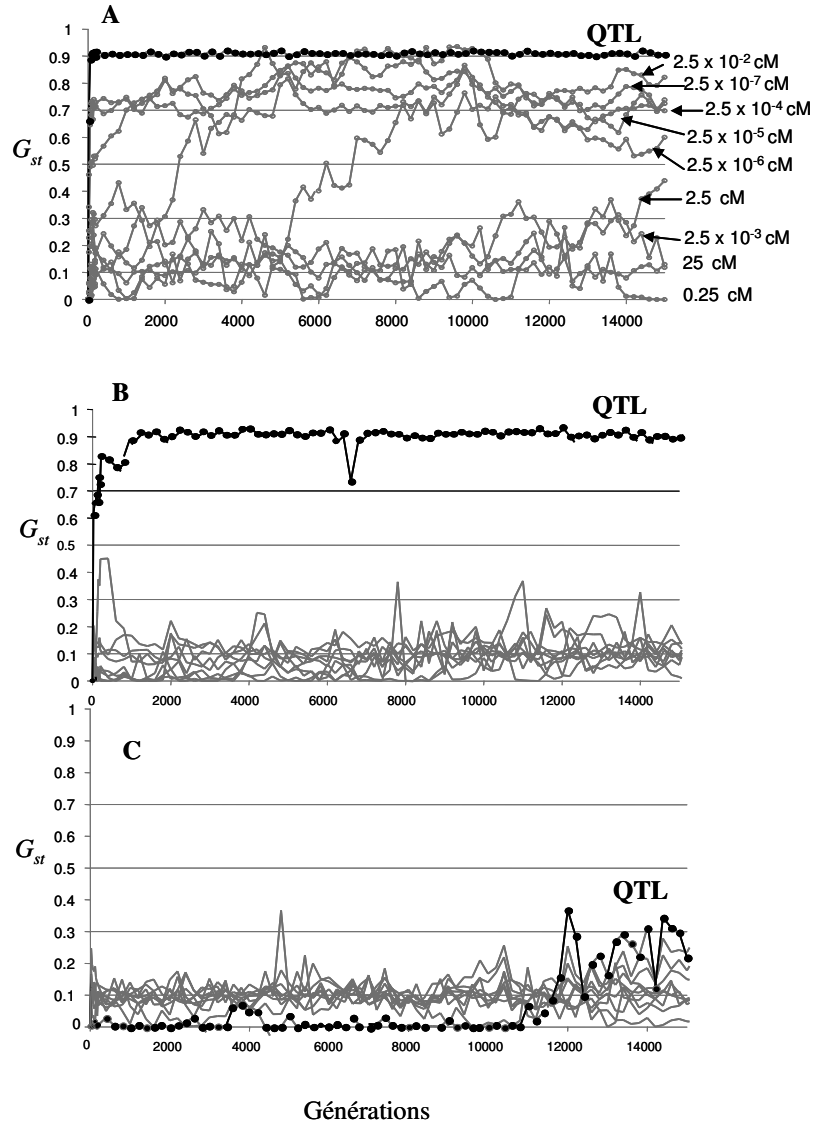


Figure 2 : Évolution de la différenciation aux QTL et aux marqueurs neutres dans différentes situations pour un scénario avec une forte sélection optimisante intrapopulation et une forte sélection diversifiante entre populations : A) valeurs de G_{st} très fortes aux QTL (en noir) et présence d'effets d'entraînement aux marqueurs neutres liés (en gris); B) valeurs de G_{st} très fortes aux QTL et absence d'effets d'entraînement; C) très faibles valeurs de G_{st} aux QTLs jusqu'à la génération 11 900 environ, et effets d'entraînement visibles quand la différenciation augmente aux QTL.

Une distance de 0,5 cM peut correspondre à une distance physique de plusieurs dizaines ou plusieurs centaines de milliers de paires de bases (Mpb), en fonction de la taille du génome des espèces (environ 6 Mpb chez le pin maritime). D'autre part, dans un scénario de sélection diversifiante tel que celui qui est modélisé, on s'attend à ce que la sélection agissant sur les fréquences alléliques aux QTL soit souvent directionnelle, en particulier dans les populations situées aux extrémités du gradient de sélection. On devrait alors observer de forts effets d'entraînement intragène, dans le cas d'un événement de sélection directionnelle sur une mutation récente qui serait avantageuse, sur les sites polymorphes proches. Par contre, dans le cas d'un événement de sélection plus ancien avec fixation de l'allèle favorable (par exemple depuis au moins 250 générations pour $N_e = 500$, d'après [33]) la puissance de détection du balayage sélectif, notamment en utilisant le test de Fay et Wu [7] ou celui de Tajima [37], diminue très rapidement et sera seulement de l'ordre de 5 % [33], [19]. Ceci pourrait expliquer les résultats précédents obtenus chez le pin maritime, où la plupart des tests de neutralité réalisés intrapopulation ne sont pas significatifs pour les gènes C4H et CAD, alors que leur différenciation est largement supérieure à celle observée sur un échantillon de marqueurs neutres. Des facteurs déterminants pour la détection d'un événement de sélection seraient donc la période depuis la fixation de la mutation favorable et l'intensité de la sélection ayant opéré sur cette mutation (ou bien la vitesse du balayage sélectif).

4. CONCLUSION, VALORISATIONS ET PERSPECTIVES D'APPLICATION À LA GESTION DE LA DIVERSITÉ ADAPTATIVE

Une grande variation des patrons de diversité et de différenciation nucléotidiques, et des écarts à la neutralité ont été observés entre gènes impliqués dans la formation des parois des tissus ligneux. Cette variation a aussi été observée dans les résultats obtenus par simulation. Pour la partie expérimentale, les écarts à la neutralité mis en évidence pour certains gènes révèlent une signature moléculaire de la sélection sur des caractères liés à l'adaptation, qui ne peuvent pas s'expliquer par des effets démographiques. On a vu aussi l'intérêt d'estimer la différenciation nucléotidique entre populations naturelles, pour identifier des mutations particulières dans le génome s'écartant des patrons de diversité neutre. Ce type d'étude s'inscrit dans des approches de génomique des populations

qui sont en plein essor sur des espèces modèles, et qui visent à identifier des zones potentiellement affectées par la sélection en parcourant systématiquement le génome grâce à un grand nombre de marqueurs cartographiés [24]. L'intérêt de ce type d'approche pour des espèces qui ne sont pas des modèles dépendra de l'importance des fenêtres de déséquilibres de liaison le long des chromosomes.

Sur l'ensemble des SNPs détectés, plus de la moitié (94) ont été ajoutés à la base de données db-SNP (<http://www.ncbi.nlm.nih.gov/SNP/>), et ont servi de lot de référence pour la mise au point de paramètres d'entrée d'une série de programmes pour la détection automatique de SNPs dans la banque d'ESTs de pin maritime [22].

Ces résultats pourront être valorisés en combinant les informations de niveaux de polymorphisme et de différenciation, et celles issues des tests de neutralité, pour le développement de critères de choix de sites polymorphes à valider ultérieurement dans des tests d'associations avec la variation quantitative. Une première application a été réalisée pour les teneurs en lignine, en sucres (cellulose, hémicelluloses) dans une population d'amélioration à base large de pin maritime [5]. Les associations remarquables démontrées pour quelques SNPs (plus de 15 % de la variation phénotypique des caractères expliquée) valident le rôle fonctionnel des gènes correspondants. Une fois ce rôle établi, les marqueurs SNPs identifiés peuvent servir d'outils moléculaires, soit comme critère indirect pour la sélection précoce de certains génotypes dans les populations d'amélioration, soit pour l'évaluation et la conservation de la diversité des caractères adaptatifs auxquels ils sont associés en populations naturelles. Ce dernier point mérite d'être approfondi, notamment par des études d'associations *in situ*.

REMERCIEMENTS

Nous remercions le Bureau des Ressources Génétiques et la Direction de l'espace rural et de la forêt (Contrat INRA n° B03846), ainsi que l'Union européenne (GEMINI, QLRT-1999-00942), et la région Aquitaine (CCRRDT 20000307007) pour leur soutien financier.

RÉFÉRENCES

- [1] Charlesworth B., Nordborg M., Charlesworth D., The effects of local selection, balanced polymorphism, and background selection on equilibrium patterns of genetic diversity in subdivided populations, *Genet. Res.* 70 (1997) 155-174.
- [2] Danjon F., Stand features and height growth in a 36-year-old maritime pine provenance test, *Silv. Genet.* 43(1) (1994) 52-62.
- [3] Derory J., Mariette S., Gonzalez-Martinez S.C., Chagné D., Madur D., Gerber S., Ribeiro M.M., Plomion C., What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenances certification strategies?, *Ann. For. Sci.* 59 (2002) 699-708.
- [4] Dvornyk V., Sirviö A., Mikkonen M., Savolainen O., Low Nucleotide Diversity at the *pal1* Locus in the Widely Distributed *Pinus sylvestris*, *Mol. Biol. Evol.* 19 (2002) 179-188.
- [5] Eveno E, Garnier-Géré P, Rodrigez J-C, Da Silva Perrez D, Plomion C., Association study between candidate genes of wood formation and wood properties in a conifer species, en préparation.
- [6] Excoffier L. Smouse P.E., Quattro J.M., Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, *Genetics* 131 (1992) 479-491.
- [7] Fay J.C., Wu C.-I., Hitchhiking under positive Darwinian selection, *Genetics* 155 (2000) 1405-1413.
- [8] Ford M.J., Applications of selective neutrality tests to molecular ecology, *Molecular Ecology* 11 (2002) 1245-1262.
- [9] Fu Y.-X., Statistical tests of neutrality of mutations against population growth, hitch-hiking and background selection, *Genetics* 147 (1997) 915-925.
- [10] Garnier-Géré P., Austerlitz F., Le Corre V., Kremer A., Multilocus patterns of adaptive hitch-hiking: I- effects on differentiation between populations of allogamous species, en préparation.
- [11] Garnier-Géré P.H., Bedon F., Austerlitz F., Pot D., Kremer A., Plomion C., Adaptive evolution of genes involved in wood formation in maritime pine populations (*Pinus pinaster* Ait.). En préparation.
- [12] Godelle B., Austerlitz F., Brachet S., Colas B., Cuguen J., Gandon S., Gouyon P.-H., Lefranc M., Olivieri I., Reboud X., Vitalis R., Système génétique, polymorphisme neutre et sélectionné : implications en biologie de la conservation, *Génét. Sél. évol.* 30 (1998) 515-528.
- [13] Halushka M.K., Fan J.B., Bentley K., Hsie L., Shen N., Weder A., Cooper R., Lipshutz R., Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis, *Nature Genet.* 22(3) (1999) 239-247.
- [14] Hamrick J.L., Godt M.J.W., Sherman-Broyles S.L., Factors influencing levels of genetic diversity in woody plant species, *New Forests* 6 (1992) 95-124.

- [15] Hudson R.R., Kreitman M., Aguadé M., A test of neutral molecular evolution based on nucleotide data, *Genetics* 116 (1987) 153-159.
- [16] Hudson R.R., Linkage disequilibrium and recombination, in Balding DJ, Bishop M, Cannings C. (eds), *Handbook of statistical Genetics*, Wiley & Sons, England, 2001, pp. 309-324.
- [17] Hudson R.R., Two-locus sampling distributions and their application, *Genetics* 159 (2001) 1805-1817.
- [18] Innan H., Nordborg M., The extent of Linkage disequilibrium and haplotype sharing around a polymorphic site, *Genetics* 165 (2003) 437-444.
- [19] Kim Y., Stephan W., Detecting a local signature of genetic hitchhiking along a recombining chromosome, *Genetics* 160 (2002) 765-777.
- [20] Kreitman M., Methods to detect selection in natural populations with applications to the human, *Ann. Rev. Genom. Genet.* 1 (2000) 539-559.
- [21] Le Corre V., Kremer A., Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection, *Genetics* 164 (2003) 1205-1219.
- [22] Le Dantec L., Chagné D., Pot D., Cantin O., Garnier-Géré P., Bedon F., Frigerio J.-M., Chaumeil P., Léger P., Garcia V., Laignet F., de Daruvar A., Plomion C. Automated SNP detection in expressed sequence tags: statistical consideration and application in maritime pine, *Plant Mol. Biol.* (2004), sous presse.
- [23] Le Provost G., Paiva J., Pot D., Brach J., Plomion C., Seasonal variation in transcript accumulation in wood forming tissues of maritime pine (*Pinus pinaster* Ait.) with emphasis on a cell wall Glycine Rich Protein, *Planta* 217 (2003) 820-830.
- [24] Luikart G., England P.R., Tallmon D., Jordan S., Taberlet P. The power and promise of population genomics: from genotyping to genome typing, *Nat. Reviews Genetics* 4 (2003) 981-994.
- [25] Machon N., Baradat D., Godelle B., *Metapop* a program simulating evolutionary processes acting on metapopulations, Paris, France: Laboratoire ES, Paris-Sud, 1995.
- [26] Mariette S., Chagné D., Lézier C., Pastuska P., Raffin A., Plomion C., Kremer A., Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers, *Heredity* 86 (2001) 469-479.
- [27] Maynard Smith J., Haigh J., The hitchhiking effect of a favourable gene, *Genet. Res.* 23 (1974) 23-35.
- [28] Nei M., Kumar S., *Molecular evolution and phylogenetics*, Oxford University Press, New-York, 2000.
- [29] Nei M., Li W.-H., Mathematical model for studying genetic variation in terms of restriction endonuclease, *Proc. Natl. Acad. Sci. USA* 76 (1979) 5269-5273.
- [30] Nei M., *Molecular Evolutionary Genetics*, Columbia University Press, New-York, 1987, 512 p.
- [31] Nordborg M., Tavaré S., Linkage disequilibrium: what history has to tell us?, *Trends in genetics* 18(2) (2002) 83-90.

Déséquilibre de liaison et cartographie fine

- [32] Paquerau M.M., Flores et climats post-glaciaires en Gironde, Actes de la Société linnéenne de Bordeaux 101 (1964) 1-156
- [33] Przeworski M., The signature of positive selection at randomly chosen loci, *Genetics* 160 (2002) 1179-1189.
- [34] Rozas J., Rozas R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis, *Bioinformatics* 15 (1999) 174-175.
- [35] Sato Y., Bao W.L., Sederoff R., Whetten R., Molecular cloning and expression of eight laccase cDNAs in loblolly pine (*Pinus taeda*), *Journal of Plant Research*, 114 (2001) 147-155.
- [36] Slatkin M., Wiehe T., Genetic hitchhiking in a subdivided population, *Genet. Res.* 71 (1998) 155-160.
- [37] Tajima F., Statistical method for testing the neutral mutation hypothesis by DNA polymorphism, *Genetics* 123 (1989) 585-595.
- [38] Tenaillon M.I., Sawkins M.C., Long A.D., Gaut R.L., Doebley J.F., Gaut B.S., Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *Mays* L.), *Proc. Nat. Acad. Science USA.* vol. 98(16) (2001) 9161-9166.
- [39] Watterson G., On the number of segregating sites in genetical models without recombination, *Theor. Popul. Biol.* 7 (1975) 256-276.
- [40] Weir B.S. *Genetic data analysis II : Methods for discrete population genetic data*, Sinauer Assoc. Inc., Sunderland, MA, USA, 1996, 445 p.
- [41] Zhang X.H., Chiang V.L., Molecular cloning of 4-coumarate: coenzyme A ligase in loblolly pine and the role of this enzyme in the biosynthesis of lignin in compression wood, *Plant Physiol.* 108 (1997) 65-74.
- [42] Zhu Y.L., Song Q.J., Hyten D.L., Van Tassell C.P., Matukumalli L.K., Grimm D.R., Hyatt S.M., Fickus E.W., Young N.D., Cregan P.B., Single-Nucleotide Polymorphisms in Soybean, *Genetics* 163 (2003) 1123-1134.